# Further Statistic
# Topical
# 2014-2022

# Table of Contents

**1. Continuous Random Variables**
**2. Inferential Statistics**
**3. Chi-Squared Tests**
**4. Non-Parametric Tests**
**5. Probability Generating Functions**

# Continuous Random Variable

**Q1.**

**3** The continuous random variable $X$ has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} kx(4-x) & 0 \leqslant x < 2, \\ k(6-x) & 2 \leqslant x \leqslant 6, \\ 0 & \text{otherwise,} \end{cases}$$

where $k$ is a constant.

**(a)** Show that $k = \frac{3}{40}$. [1]

.....

.....

.....

.....

.....

**(b)** Given that $\mathrm{E}(X) = 2.5$, find $\mathrm{Var}(X)$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(c)** Find the median value of $X$. [4]

**Q2.**

**4** The continuous random variable $X$ has probability density function f given by

$$f(x) = \begin{cases} \frac{3}{8}\left(1 + \frac{1}{x^2}\right) & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

**(a)** Find $E\left(\sqrt{X}\right)$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable $Y$ is given by $Y = X^2$.

**(b)** Find the probability density function of $Y$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(c)** Find the 40th percentile of $Y$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q3.**

**3** The continuous random variable $X$ has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{81}x^2 & 0 \leqslant x \leqslant 9, \\ 1 & x > 9. \end{cases}$$

**(a)** Find $\mathrm{E}\left(\sqrt{X}\right)$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Find $\mathrm{Var}\left(\sqrt{X}\right)$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

**(c)** The random variable $Y$ is given by $Y^3 = X$. Find the probability density function of $Y$. [3]

**Q4.**

**6** The continuous random variable $X$ has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{8} & 0 \leqslant x < 1, \\ \frac{1}{28}(8-x) & 1 \leqslant x \leqslant 8, \\ 0 & \text{otherwise.} \end{cases}$$

**(a)** Find the cumulative distribution function of $X$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Find the value of the constant $a$ such that $\mathrm{P}(X \leqslant a) = \frac{5}{7}$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable $Y$ is given by $Y = \sqrt[3]{X}$.

**(c)** Find the probability density function of $Y$. [5]

**Q5.**

**2** The continuous random variable $X$ has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < -1, \\ \frac{1}{2}(1+x)^2 & -1 \leqslant x \leqslant 0, \\ 1 - \frac{1}{2}(1-x)^2 & 0 < x \leqslant 1, \\ 1 & x > 1. \end{cases}$$

**(a)** Find the probability density function of $X$. [2]

**(b)** Find $\mathrm{P}\left(-\frac{1}{2} \leqslant X \leqslant \frac{1}{2}\right)$. [2]

**(c)** Find $\mathrm{E}(X^2)$. [2]

**(d)** Find $\mathrm{Var}(X^2)$. [2]

**Q6.**

**3** The continuous random variable $X$ has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} a + \frac{1}{5}x & 0 \leqslant x < 1, \\ 2a - \frac{1}{5}x & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where $a$ is a constant.

**(a)** Find the value of $a$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Find $\mathrm{E}(X^2)$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(c)** Find the cumulative distribution function of $X$. [3]

**Q7.**

**3** The continuous random variable $X$ has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{3}{16}\left(2 - \sqrt{x}\right) & 0 \leqslant x < 1, \\ \dfrac{3}{16\sqrt{x}} & 1 \leqslant x \leqslant 9, \\ 0 & \text{otherwise.} \end{cases}$$

**(a)** Find E($X$). [3]

The random variable $Y$ is such that $Y = \sqrt{X}$.

**(b)** Find the probability density function of $Y$. [5]

**Q8.**

**3** The continuous random variable $X$ has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{5}x & 0 \leqslant x < 2, \\ \frac{2}{15}(5-x) & 2 \leqslant x \leqslant 5, \\ 0 & \text{otherwise.} \end{cases}$$

**(a)** Find the cumulative distribution function of $X$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Find the median value of $X$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2020

**(c)** Find $\mathrm{E}(X^2)$. [2]

**(d)** Find $\mathrm{P}(1 \leqslant X \leqslant 3)$. [2]

**Q9.**

**6** The continuous random variable $X$ has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{60}(16x - x^2) & 0 \leqslant x \leqslant 6, \\ 1 & x > 6. \end{cases}$$

**(a)** Find the interquartile range of $X$. [4]

**(b)** Find $\mathrm{E}(X^3)$. [4]

The random variable $Y$ is such that $Y = \sqrt{X}$.

**(c)** Find the probability density function of $Y$. [3]

**Q10.**

**4** The continuous random variable $X$ has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 2, \\ \frac{1}{60}x^2 - \frac{1}{15} & 2 \leqslant x \leqslant 8, \\ 1 & x > 8. \end{cases}$$

**(a)** Find $\mathrm{P}(3 \leqslant X \leqslant 6)$. [1]

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Find $\mathrm{E}\left(\sqrt{X}\right)$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(c)** Find $\mathrm{Var}\left(\sqrt{X}\right)$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(d)** The random variable $Y$ is defined by $Y = X^3$. Find the probability density function of $Y$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q11.**

**7** The continuous random variable $X$ has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \dfrac{3}{4x^2} + \dfrac{1}{4} & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

**(i)** Find the distribution function of $X$. [3]

**(ii)** Find the exact value of the interquartile range of $X$. [5]

**Q12.**

**7** The time, $T$ days, before an electrical component develops a fault has distribution function F given by

$$\mathrm{F}(t) = \begin{cases} 1 - \mathrm{e}^{-at} & t \geqslant 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $a$ is a positive constant. The mean value of $T$ is 200.

**(i)** Write down the value of $a$. [1]

.....

.....

**(ii)** Find the probability that an electrical component of this type develops a fault in less than 150 days. [2]

.....

.....

.....

.....

.....

A piece of equipment contains $n$ of these components, which develop faults independently of each other. The probability that, after 150 days, at least one of the $n$ components has not developed a fault is greater than 0.99.

**(iii)** Find the smallest possible value of $n$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q13.**

**10** The random variable $X$ has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{30}\left(\frac{8}{x^2} + 3x^2 - 14\right) & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

**(i)** Find the distribution function of $X$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable $Y$ is defined by $Y = X^2$.

**(ii)** Find the probability density function of $Y$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2019

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(iii)** Find the value of $y$ such that $P(Y < y) = 0.8$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q14.**

**6** The continuous random variable $X$ has distribution function given by

$$\mathrm{F}(x) = \begin{cases} 1 - \mathrm{e}^{-0.4x} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

**(i)** Find $\mathrm{P}(X > 2)$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** Find the interquartile range of $X$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q15.**

**9** The continuous random variable $X$ has probability density function given by

$$f(x) = \begin{cases} \frac{1}{20}\left(3 - \frac{1}{\sqrt{x}}\right) & 1 \leqslant x \leqslant 9, \\ 0 & \text{otherwise.} \end{cases}$$

The random variable $Y$ is defined by $Y = \sqrt{X}$.

**(i)** Show that the probability density function of $Y$ is given by

$$g(y) = \begin{cases} \frac{1}{10}(3y - 1) & 1 \leqslant y \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$ [7]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** Find the mean value of $Y$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q16.**

**6** The continuous random variable $X$ has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{80}\left(3\sqrt{x} - \frac{8}{\sqrt{x}}\right) & 4 \leqslant x \leqslant 16, \\ 0 & \text{otherwise.} \end{cases}$$

**(i)** Find the distribution function of $X$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable $Y$ is defined by $Y = \sqrt{X}$.

**(ii)** Find the probability density function of $Y$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2018

**Q17.**

**7** The continuous random variable $X$ has distribution function given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{90}(x^2 + x^4) & 0 \leqslant x \leqslant 3, \\ 1 & x > 3. \end{cases}$$

The random variable $Y$ is defined by $Y = X^2$.

**(i)** Find the probability density function of $Y$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** Find the mean value of $Y$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q18.**

**8** The continuous random variable $X$ has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{4}(x-1) & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

**(i)** Find the distribution function of $X$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable $Y$ is defined by $Y = (X-1)^3$.

**(ii)** Find the probability density function of $Y$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2017

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(iii)** Find the median value of $Y$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q19.**

**9** The continuous random variable $X$ has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} 0 & x < 0, \\ a\mathrm{e}^{-x\ln 2} & x \geqslant 0, \end{cases}$$

where $a$ is a positive constant.

**(i)** Find the value of $a$. [2]

.....

.....

.....

.....

**(ii)** State the value of $\mathrm{E}(X)$. [1]

.....

.....

**(iii)** Find the interquartile range of $X$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2017

The variable $Y$ is related to $X$ by $Y = 2^X$.

**(iv)** Find the probability density function of $Y$. [5]

**Q20.**

**7** The random variable $X$ has probability density function f given by

$$f(x) = \begin{cases} 0.2e^{-0.2x} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

**(i)** Find the distribution function of $X$. [2]

.....

.....

.....

.....

.....

.....

**(ii)** Find $P(X > 2)$. [2]

.....

.....

.....

.....

.....

**(iii)** Find the median of $X$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q21.**

**4**

**8** The random variable $X$ has probability density function f given by

$$f(x) = \begin{cases} 2e^{-2x} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

**(i)** Find the distribution function of $X$. [2]

**(ii)** Find the median value of $X$. [3]

The random variable $Y$ is defined by $Y = e^X$.

**(iii)** Find the probability density function of $Y$. [4]

© UCLES 2016

**Q22.**

**7** The lifetime, in months, of a particular type of light bulb is a random variable $T$ and its probability density function, f, is given by

$$\mathrm{f}(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda t} & t \geqslant 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda$ is a positive constant. The manager of a shopping mall uses this type of light bulb and he notes that, out of 1000 new light bulbs, 280 were still working after 12 months.

**(i)** Find an estimate for the value of $\lambda$. [3]

**(ii)** Hence estimate the mean of $T$. [1]

An improved version of the light bulb is produced. The lifetime of the improved version also has a negative exponential distribution, with a mean 25% larger than the original mean.

**(iii)** What percentage of the improved version of the light bulb fail within the first 12 months? [4]

© UCLES 2016

**[Turn over**

**Q23.**

**7** The random variable $X$ has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{6}x & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

**(i)** Find the distribution function of $X$. [3]

The random variable $Y$ is defined by $Y = X^3$. Find

**(ii)** the probability density function of $Y$, [2]

**(iii)** the value of $k$ for which $P(Y \geqslant k) = \frac{7}{12}$. [3]

© UCLES 2016

**Q24.**

**9** The continuous random variable $X$ has probability density function given by

$$f(x) = \begin{cases} 0 & x < 2, \\ ae^{-(x-2)} & x \geqslant 2, \end{cases}$$

where $a$ is a constant. Show that $a = 1$. [1]

Find the distribution function of $X$ and hence find the median value of $X$. [5]

The random variable $Y$ is defined by $Y = e^X$. Find

**(i)** the probability density function of $Y$, [4]

**(ii)** $P(Y > 10)$. [2]

© UCLES 2015

**Q25.**

**9** The continuous random variable $X$ has distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 2, \\ \frac{1}{8}x - \frac{1}{4} & 2 \leqslant x \leqslant 10, \\ 1 & x > 10. \end{cases}$$

Find the value of $k$ for which $\mathrm{P}(X \geqslant k) = 0.6$. [3]

The random variable $Y$ is defined by $Y = 2\ln X$. Find the distribution function of $Y$. [3]

Find the probability density function of $Y$ and sketch its graph. [4]

© UCLES 2014

# MS
# Continuous Random Variable

**Q1.** May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(a) | $k2x^2 - \frac{1}{3}x^3 + k\left[6x - \frac{1}{2}x^2\right] = 1$<br>$k\left[8 - \frac{8}{3} + 36 - 18 - 12 + 2\right] = \frac{40}{3}k = 1$<br>$k = \frac{3}{40}$ | **B1** | Integrate and equate to 1, no errors seen. |
| | | **1** | |
| 3(b) | $\mathrm{E}(X^2) = k\int_0^2 \left(4x^3 - x^4\right)dx + k\int_2^6 \left(6x^2 - x^3\right)\mathrm{d}x =$<br>$k\left[x^4 - \frac{1}{5}x^5\right] + k\left[2x^3 - \frac{1}{4}x^4\right] = k\left[9.6 + 96\right] = 7.92$ | **M1*** | Attempt at $\mathrm{E}(X^2)$, correct limits, integrated. |
| | $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) =$ *their* $(7.92) - 2.5^2$ | **depM1** | Use of correct formula, with their $\mathrm{E}(X^2)$. |
| | 1.67 | **A1** | |
| | | **3** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(c) | $\int_0^2 kx(4-x)\,dx + \int_2^m k(6-x) = \frac{1}{2}$ | **M1*** | Correct expression and attempt to integrate. |
| | $\int_0^2 kx(4-x)\,dx = \frac{2}{5}$ | **B1** | May be implied |
| | $\frac{2}{5} + k\left[6x - \frac{1}{2}x^2\right] = \frac{1}{2}$<br>$6m - \frac{1}{2}m^2 - 12 + 2 = \frac{1}{10k}$<br>$m^2 - 12m + \frac{68}{3} = 0$ | **depM1** | First integral and $\frac{1}{2}$ may be seen combined as $\frac{1}{10}$.<br><br>Obtain quadratic equation. |
| | $m = 6 \pm \frac{2}{3}\sqrt{30}$<br>$m = 6 - \frac{2}{3}\sqrt{30}\,(= 2.35)$ | **A1** | Single answer. |
| | | **4** | |

**Q2.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4(a) | $\mathrm{E}(\sqrt{X}) = \int \sqrt{x}\,\frac{3}{8}\left(1 + \frac{1}{x^2}\right)\mathrm{d}x \left(= \int_1^3 \frac{3}{8}\left(x^{0.5} + x^{-1.5}\right)\mathrm{d}x\right)$ | **M1** | |
| | $\frac{3}{8}\left[\frac{2}{3}x^{1.5} - 2x^{-0.5}\right]$ | **A1** | |
| | $\frac{3}{8}\left(2\sqrt{3} - \frac{2}{\sqrt{3}} - \frac{2}{3} + 2\right) = \frac{1}{2}\left(1 + \sqrt{3}\right)$ $(= 1.37)$ | **A1** | |
| | | **3** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4(b) | $F(x)=\frac{3}{8}\left(x-\frac{1}{x}\right)$ | **M1** | Integrate. |
| | $G(y)=\frac{3}{8}\left(\sqrt{y}-\frac{1}{\sqrt{y}}\right)$ | **M1** | Change variable throughout in their CDF. |
| | $g(y)=\frac{3}{16}\left(y^{-0.5}+y^{-1.5}\right), \quad (1\leqslant y\leqslant 9)$ | **A1** | Differentiate to find g($y$). |
| | 0 otherwise | **A1** | $1\leqslant y\leqslant 9$ and 0 otherwise, dependent on second M1. |
| | | **4** | |
| 4(c) | $F(p)=0.4$ so $\frac{3}{8}\left(p-\frac{1}{p}\right)=0.4$ | **M1*** | Correct method. |
| | $15p^2-16p-15=0$<br>$(5p+3)(3p-5)=0, \quad p=\frac{5}{3}$ | **M1 dep** | Obtain quadratic and solve. |
| | 40th percentile of $Y$ is $\frac{25}{9}$ | **A1** | |
| | | **3** | |

**Q3.** May/June 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(a) | $\int_0^9 \frac{2}{81}x^{1.5}\,dx$ | **B1** | $f(x)=\frac{2}{81}x$, correct expression as integrand. |
| | $\frac{2}{81}\times\frac{2}{5}x^{2.5}$ | **M1** | Integrate, FT only on their PDF expression. |
| | $\frac{12}{5}=2.4$ | **A1** | |
| | | **3** | |
| 3(b) | $\int_0^9 \frac{2}{81}x^2\,dx-(their\ a)^2$ | **M1** | |
| | $\frac{2}{81}\times\frac{9^3}{3}-2.4^2=\frac{6}{25}$ | **A1** | |
| | | **2** | |
| 3(c) | $G(y)=\frac{1}{81}y^6$ | **M1** | |
| | $g(y)=\frac{2}{27}y^5$ | **M1** | |
| | Fully correct including 'for $0\leqslant y\leqslant\sqrt[3]{9}$ and 0 otherwise' | **A1** | Condone 2.08 |
| | | **3** | |

**Q4.** May/June 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(a) | $F(x)=\begin{cases}\frac{1}{8}x & 0\leqslant x<1,\\ \frac{1}{28}\left(8x-\frac{1}{2}x^2\right)-\frac{1}{7} & 1\leqslant x\leqslant 8.\end{cases}$ | **M1**<br>**A1** | Integrate both parts (all powers of $x$ increase by 1), allow M1 if no constant term.<br>Both parts correct, with domains.<br>Any equivalent form, for example: $\frac{2}{7}x-\frac{1}{56}x^2-\frac{1}{7}$. |
| | $F(x)=0$ for $x<0$ and $F(x)=1$ for $x>8$. | **A1** | |
| | **Alternative method for question 6(a)** | | |
| | $F(x)=\begin{cases}\frac{1}{8}x & 0\leqslant x<1,\\ 1-\frac{1}{56}(8-x)^2 & 1\leqslant x\leqslant 8.\end{cases}$ | **M1**<br>**A1** | Integrate both parts (all powers of $x$ increase by 1), allow M1 if no constant term.<br>Both parts correct, with domains.<br>Any equivalent form, for example: $\frac{1}{56}\left(16x-x^2-8\right)$. |
| | $F(x)=0$ for $x<0$ and $F(x)=1$ for $x>8$. | **A1** | |
| | | **3** | |
| 6(b) | $F(a)=\frac{5}{7}$ leading to $\frac{1}{28}\left(8a-\frac{1}{2}a^2\right)-\frac{1}{7}=\frac{5}{7}$ | ***M1** | Correct method and attempt to simplify, ft their part **(a).** |
| | $a^2-16a+48=0$<br>$a=4$ $[a=12]$ | **DM1** | Obtain quadratic and solve. |
| | $a=4$ | **A1** | Correct single answer, WWW. |
| | | **3** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(c) | $\mathrm{G}(y)=\begin{cases}\frac{1}{8}y^3 & 0 \leqslant y<1,\\ \frac{1}{56}\left(16y^3-y^6\right)-\frac{1}{7} & 1 \leqslant y \leqslant 2.\end{cases}$ | **M1**<br>**A1** | Substitute $x = y^3$ into both parts of their F($x$) from part **(a)**.<br>Correct functions in terms of $y$, any equivalent form. |
| | $0 \leqslant y<1$ and $1 \leqslant y \leqslant 2$seen in CDF or PDF | **A1** | Condone $\leqslant$ instead of $<$. |
| | $\mathrm{g}(y)=\begin{cases}\frac{3}{8}y^2 & 0 \leqslant y<1,\\ \frac{3}{28}\left(8y^2-y^5\right) & 1 \leqslant y<2,\\ 0 & \text{otherwise.}\end{cases}$ | **M1**<br>**A1** | Differentiate their G($y$), all powers of $y$ decrease by 1.<br>Must include 'otherwise'. |
| | | **5** | |

| | | | |
|---|---|---|---|
| 6(c) | **Alternative method for question 6(c)** | | |
| | $y=x^{\frac{1}{3}},\ \mathrm{d}y=\frac{1}{3}x^{-\frac{2}{3}}\mathrm{d}x,\ \mathrm{d}x=3y^2\mathrm{d}y$ | **M1** | |
| | $0 \leqslant x<1,\ \mathrm{f}(x)\mathrm{d}x=\frac{1}{8}\mathrm{d}x=\frac{3}{8}y^2\mathrm{d}y$ | **M1** | |
| | $\mathrm{g}(y)=\frac{3}{8}y^2$ | **A1** | |
| | $1 \leqslant x \leqslant 8,\ \mathrm{f}(x)\mathrm{d}x=\frac{1}{28}(8-x)\mathrm{d}x=\frac{1}{28}\left(8-y^3\right)3y^2\mathrm{d}y$<br>$g(y)=\frac{3}{28}\left(8y^2-y^5\right)$ | **A1** | |
| | $0 \leqslant y<1$ and $1 \leqslant y \leqslant 2$ seen in CDF or PDF | **A1** | |
| | | **5** | |

**Q5.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2(a) | $f(x)=\begin{cases}1+x & -1\leqslant x\leqslant 0,\\ 1-x & 0<x\leqslant 1,\\ 0 & \text{otherwise.}\end{cases}$ | **M1** | Differentiation attempted. |
| | | **A1** | All correct, including 0 otherwise. |
| | | **2** | |
| 2(b) | $P\left(-\frac{1}{2}\leqslant X\leqslant\frac{1}{2}\right)=F\left(\frac{1}{2}\right)-F\left(-\frac{1}{2}\right)=\frac{7}{8}-\frac{1}{8}$ | **M1** | Can do by integration with correct limits. |
| | $\frac{3}{4}$ | **A1** | |
| | | **2** | |
| 2(c) | $\int_{-1}^{0}x^2(1+x)\,dx+\int_{0}^{1}x^2(1-x)\,dx$<br>$\left[\frac{1}{3}x^3+\frac{1}{4}x^4\right]+\left[\frac{1}{3}x^3-\frac{1}{4}x^4\right]$ | **M1** | |
| | $\frac{1}{12}+\frac{1}{12}=\frac{1}{6}$ | **A1** | |
| | | **2** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2(d) | $\int_{-1}^{0}x^4(1+x)\,dx+\int_{0}^{1}x^4(-x)\,dx-\left(\frac{1}{6}\right)^2$<br>$\left[\frac{1}{5}x^5+\frac{1}{6}x^6\right]+\left[\frac{1}{5}x^5-\frac{1}{6}x^6\right]-\frac{1}{36}$ | **M1** | Complete method with their mean squared explicit. |
| | $\frac{1}{15}-\frac{1}{36}=\frac{7}{180}$ | **A1** | |
| | | **2** | |

**Q6.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(a) | Total prob = 1, so $\int_0^1\left(a+\frac{1}{5}x\right)dx+\int_1^2\left(2a-\frac{1}{5}x\right)dx=1$<br>$\left[ax+\frac{1}{10}x^2\right]+\left[2ax-\frac{1}{10}x^2\right]=1$ | ***M1** | Correct expression integrated and equated to 1.<br>Powers of $x$ correct. |
| | $3a-\frac{1}{5}=1$ | **DM1** | Limits used and attempt to solve. |
| | Solve to give $a=\frac{2}{5}$ | **A1** | |
| | | **3** | |
| 3(b) | $\int_0^1\frac{1}{5}(2+x)x^2\,dx+\int_1^2\frac{1}{5}(4-x)x^2\,dx$<br>$\frac{1}{5}\left[\frac{2}{3}x^3+\frac{1}{4}x^4\right]+\frac{1}{5}\left[\frac{4}{3}x^3-\frac{1}{4}x^4\right]$ | **M1** | Correct expressions integrated, FT their $a$. |
| | $\frac{13}{10}$ | **A1** | $\frac{11}{60}+\frac{67}{60}$ |
| | | **2** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(c) | $F(x)=\begin{cases}0 & x<0\\ \frac{1}{5}\left(2x+\frac{1}{2}x^2\right) & 0\leqslant x<1,\\ \frac{1}{5}\left(4x-\frac{1}{2}x^2-1\right) & 1\leqslant x\leqslant 2,\\ 1 & x>2.\end{cases}$ | **M1** | Integration of their PDF. |
| | | **A1** | Middle 2 parts correct. |
| | $=0\quad(x<0)$ and $=1\ (x>2)$ | **A1** | First and last parts correct and all domains correct, with no gaps. |
| | | **3** | |

**Q7.** May/June 2020

| Question | Answer | Marks |
|---|---|---|
| 3(a) | $\mathrm{E}(X) = \int_0^1 \frac{3}{16}x\left(2-\sqrt{x}\right)\mathrm{d}x + \int_1^9 \frac{\frac{3}{16}x}{\sqrt{x}}\mathrm{d}x$ | **M1** |
| | $\frac{3}{16}\left[x^2 - \frac{2}{5}x^{\frac{5}{2}}\right] + \frac{3}{16}\left[\frac{2}{3}x^{\frac{3}{2}}\right]$ | **A1** |
| | $\frac{269}{80}$ | **A1** |
| | | **3** |
| 3(b) | $\mathrm{F}(x) = \begin{cases} \frac{3}{8}\left(x - \frac{1}{3}x^{3/2}\right), & 0 \leqslant x < 1 \\ \frac{3}{8}x^{1/2} - \frac{1}{8}, & 1 \leqslant x \leqslant 9 \end{cases}$ | **M1A1** |
| | $\mathrm{G}(y) = \begin{cases} \frac{3}{8}\left(y^2 - \frac{1}{3}y^3\right), & 0 \leqslant y < 1 \\ \frac{3}{8}y - \frac{1}{8}, & 1 \leqslant y \leqslant 3 \end{cases}$ | **M1A1** |
| | (G($y$) = 0 for $y \leqslant 0$ and = 1 for $y \geqslant 3$)<br>$\mathrm{g}(y) = \begin{cases} \frac{3}{8}\left(2y - y^2\right), & 0 \leqslant y < 1 \\ \frac{3}{8}, & 1 \leqslant y \leqslant 3 \end{cases}$<br>= 0 otherwise | **A1** |
| | | **5** |

**Q8.**

| Question | Answer | Marks |
|---|---|---|
| 3(a) | $\mathrm{F}(x)=\begin{cases}\dfrac{x^2}{10}, & 0\leqslant x<2\\ \dfrac{1}{15}\left(10x-x^2\right)-\dfrac{2}{3}, & 2\leqslant x\leqslant 5\end{cases}$ | **M1A1** |
| | = 0 for $x < 0$ and = 1 for $x > 5$ | **A1** |
| | | **3** |
| 3(b) | F($m$) = ½ so $2m^2-20m+35=0$ | **M1** |
| | $m=5-\dfrac{1}{2}\sqrt{30}=2.26$ | **A1** |
| | | **2** |
| 3(c) | $\mathrm{E}(X^2)=\int_0^2 x^2.\dfrac{x}{5}\,\mathrm{d}x+\int_2^5\dfrac{2}{15}(5-x).x^2\,\mathrm{d}x=\left[\dfrac{x^4}{20}\right]+\left[\dfrac{2}{15}\left(\dfrac{5}{3}x^3-\dfrac{x^4}{4}\right)\right]$ | **M1** |
| | = 6.5 | **A1** |
| | | **2** |
| 3(d) | F(3) – F(1) | **M1** |
| | $=\dfrac{11}{15}-\dfrac{1}{10}=\dfrac{19}{30}$ | **A1** |
| | **Alternative method for 3(d)** | |
| | $\int_1^2\dfrac{x}{5}\,\mathrm{d}x+\int_2^3\dfrac{2}{15}(5-x)\,\mathrm{d}x=\left[\dfrac{x^2}{10}\right]+\left[\dfrac{2}{15}\left(5x-\dfrac{x^2}{2}\right)\right]$ | **M1** |
| | $=\dfrac{3}{10}+\dfrac{1}{3}=\dfrac{19}{30}$ | **A1** |
| | | **2** |

**Q9.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(a) | UQ: F(U) = 0.75: $u^2 - 16u + 45 = 0$ | **M1** | Obtain quadratic equation for $u$ or $l$ |
| | $u = 8 - \sqrt{19}$ (= 3.64) | **A1** | Value of UQ |
| | LQ: $l^2 - 16l + 15 = 0$, $l = 1$ | **A1** | Value of LQ |
| | IQR = UQ – LQ = $7 - \sqrt{19} = 2.64$ | **A1 FT** | |
| | | **4** | |
| 6(b) | $f(x) = \begin{cases} \frac{1}{30}(8-x), & 0 \leqslant x \leqslant 6 \\ 0, & \text{otherwise} \end{cases}$ | **B1** | May be implied,<br>only need to see $f(x) = \frac{1}{30}(8-x)$ |
| | Find E($X^3$) by integration:<br>$\frac{1}{30}\int_0^6 (8x^3 - x^4)\,dx$<br>$= \left[2x^4 - \frac{1}{5}x^5\right]_0^6$ | **M1 A1** | Attempt integration (with PDF not CDF)<br>limits not required<br>Correct integrated expression, with limits |
| | = 34.56 | **A1** | |
| | | **4** | |
| 6(c) | $G(y) = \frac{1}{60}(16y^2 - y^4)$ | **M1** | CDF for $Y$ |
| | $g(y) = \frac{1}{60}(32y - 4y^3)$ for $0 \leq y \leq \sqrt{6}$ | **M1A1** | Differentiate to find PDF for $Y$<br>Correct g($y$) and correct range seen anywhere |
| | | **3** | |

**Q10.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4(a) | $F(6) - F(3) = \frac{9}{20}$ or 0.45 | **B1** | |
| | | **1** | |
| 4(b) | $f(x) = \begin{cases} \frac{1}{30}x, & 2 \le x \le 8 \\ 0, & \textit{otherwise} \end{cases}$ | **B1** | May be implied.<br>Only need to see $f(x) = \frac{1}{30}x$ |
| | $E(\sqrt{X}) = \int_2^8 \frac{1}{30}x^{\frac{3}{2}}dx = \left[\frac{1}{75}x^{\frac{5}{2}}\right]$ | **M1** | Integrated |
| | = 2.34 | **A1** | 2.338…<br>Answer of 2.34 without $\frac{1}{75}x^{\frac{5}{2}}$ scores B1B1 (2/3) |
| | | **3** | |
| 4(c) | $\text{Var}(\sqrt{X}) = [\int_2^8 \frac{1}{30}x^2dx = \left[\frac{1}{90}x^3\right]] - E\left(\sqrt{X}\right)^2$<br>$\text{Var}\left(\sqrt{X}\right) = \left[\int_2^8 \frac{1}{30}x^2dx = \left[\frac{1}{90}x^3\right]_2^8\right] - E\left(\sqrt{X}\right)^2$ | **M1** | |
| | = 5.6 – 2.338²<br>= 0.133 or 0.134 | **A1** | |
| | | **2** | |
| 4(d) | $G(y) = \frac{1}{60}y^{\frac{2}{3}} - \frac{1}{15}$ | **M1** | CDF for $Y$ |
| | $g(y) = \frac{1}{90}y^{-\frac{1}{3}}$ for $8 \le y \le 512$<br>(0 otherwise) | **M1**<br>**A1** | Differentiate to find PDF for $Y$<br>Correct $g(y)$ and correct range seen anywhere |
| | | **3** | |

**Q11.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 7(i) | $F(x) = \int f(x)\,dx = -3/(4x) + x/4$ [+ $c$] | **M1** | Find or state distribution function $F(x)$ for $1 \leqslant x \leqslant 3$ |
| | $= -3/(4x) + x/4 + ½$ *or* ¼ $(-3/x + x + 2)$ | **A1** | using $F(1) = 0$ or $F(3) = 1$ to find $c$ if necessary |
| | $F(x) = 0$ $(x < or \leqslant 1)$, $F(x) = 1$ $(x > or \geqslant 3)$ | **A1** | State $F(x)$ for other values of $x$ |
| | | **3** | |
| 7(ii) | $\int_1^Q f(x)\,dx = -3/4Q + Q/4 + ½ = ¼$ [*or* ¾] (AEF) | **M1** | Formulate equation for either quartile value $Q$ |
| | $Q^2 + [or -]\ Q - 3 = 0$ | **A1** | |
| | $Q_1 = ½ (-1 + \sqrt{13})$, $Q_3 = ½ (1 + \sqrt{13})$ | **A1 A1** | Find lower quartile $Q_1$ and upper quartile $Q_3$ |
| | $Q_3 - Q_1 = 1$ | **A1** | Find interquartile range |
| | | **5** | |

**Q12.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 7(i) | $a = 1/200$ *or* 0·005 | **B1** | State $a$ or find $a$ by equating mean value to $1/a$ |
| | | **1** | |
| 7(ii) | $p = P(T < 150) = F(150) = 1 - e^{-150a}$ | **M1** | Find $P(T < 150)$ |
| | $p = 1 - e^{-0.75} = 0{\cdot}528$ | **A1** | |
| | | **2** | |
| 7(iii) | $1 - p^n > 0{\cdot}99$ | **M1** | Formulate condition for $n$ |
| | $0{\cdot}01 > (1 - e^{-0.75})^n$ or $0{\cdot}01 > 0{\cdot}528^n$ | **A1** | |
| | $n >$ log 0·01 / log 0·528 | **M1** | Rearrange and take logs to give bound |
| | $n > 7{\cdot}20$ [or 7·21] so $n_{min} = 8$ | **A1** | Find $n_{min}$ |
| | | **4** | |

**Q13.** October/November 2019

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 10(i) | $F(x) = \int f(x)\,dx = (1/30)\,(-8/x + x^3 - 14x)$ [+ $c$] | **M1** | Find or state distribution function $F(x)$ for $2 \leqslant x \leqslant 4$ |
| | $F(x) = (1/30)\,(-8/x + x^3 - 14x + 24)$ | **M1** | Using $F(2) = 0$ or $F(4) = 1$ to find $c$ if necessary. AEF |
| | $F(x) = 0$ ($x <$ *or* $\leqslant 2$), $F(x) = 1$ ($x >$ *or* $\geqslant 4$) | **A1** | State $F(x)$ for other values of $x$ |
| | | **3** | |
| 10(ii) | $G(y) = P(Y < y) = P(X^2 < y)$<br>$G(y) = P(X < \sqrt{y}) = F(\sqrt{y})$<br>$G(y) = (1/30)\left(-8/y^{\frac{1}{2}} + y^{\frac{3}{2}} - 14y^{\frac{1}{2}} + 24\right)$ | **M1**<br><br>**A1** | Find or state $G(y)$ for $2 \leqslant x \leqslant 4$ from $Y = X^2$<br>(allow < or $\leqslant$ throughout) |
| | **Alternative method for question 10(ii)** | | |
| | Use $x = y^{\frac{1}{2}}$ to find<br>$f(x) = (1/30)(8/y + 3y - 14), \frac{dx}{dy} = -\frac{1}{2}y^{-\frac{1}{2}}$ | **(M1**<br><br>**A1)** | Find $f(x)$ *and* $\frac{dx}{dy}$ for use in $g(y) = f(x) \times \left\lvert\frac{dy}{dx}\right\rvert$ |
| | $g(y)[= G'(y)] = (1/30)\left(4/y^{\frac{3}{2}} + (3/2)y^{\frac{1}{2}} - 7/y^{\frac{1}{2}}\right)$<br>for $4 \leqslant y \leqslant 16$ [$g(y) = 0$ otherwise] | **A1**<br>**A1** | Find $g(y)$. AEF<br>State corresponding range of $y$ for $G(y)$ *or* $g(y)$ |
| | | **4** | |
| 10(iii) | $(1/30)\left(-8/y^{\frac{1}{2}} + y^{\frac{3}{2}} - 14y^{\frac{1}{2}} + 24\right) = 0.8$ | **M1** | Set $G(y) = 0{\cdot}8$ |
| | $-8 + y^2 - 14y = 0$, $y = 7 + \sqrt{57}$ *or* 14·5[5]<br>[rejecting $7 - \sqrt{57}$; allow 14·6] | **M1 A1** | Rearrange to give quadratic in $y$ and solve to find value of $y$ |
| | | **3** | |

**Q14.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(i) | $P(X > 2) = 1 - F(2) = \exp(-0{\cdot}8) = 0{\cdot}449$ | **M1 A1** | Find $P(X > 2)$. M0 for $F(2)$ |
| | | **2** | |
| 6(ii) | $F(Q) = 1 - \exp(-0{\cdot}4x) = \frac{1}{4}$ *or* $\frac{3}{4}$ | **M1** | Formulate equation for either quartile value $Q$ |
| | $Q_1 = (\ln 4/3) / (0{\cdot}4)$ [= 0·7192 ] (AEF) | **A1** | Find one [lower] quartile $Q_1$ |
| | $Q_3 = (\ln 4) / (0{\cdot}4)$ [= 3·466] (AEF) | **A1** | Find other [upper] quartile $Q_3$ |
| | $Q_3 - Q_1$ [= (ln 3) / (0·4)] = 2·75 | **A1**√ | Find interquartile range (FT on $Q_1$, $Q_3$; allow $Q_1 - Q_3$) |
| | | **4** | |

**Q15.** May/June 2018

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 9(i) | *EITHER:* $F(x) = \int f(x)\,dx = (1/20)(3x - 2\sqrt{x}\ [+ c])$<br>$c = -1$ so $F(x) = (1/20)(3x - 2\sqrt{x} - 1)$ | **M1** | Find or state distribution function $F(x)$ for $1 \leqslant x \leqslant 9$<br>(may be implied by $G(y)$) |
| | *or* $(3/20)x - (1/10)\sqrt{x} - 1/20$<br>$G(y)$ [$= P(Y < y) = P(\sqrt{X} < y) = P(X < y^2)$ ] | **A1** | Find or state $G(y)$ from $Y = \sqrt{X}$ for $1 \leqslant x \leqslant 9$ or $1 \leqslant y \leqslant 3$ |
| | $= F(y^2)$<br>$= (1/20)(3y^2 - 2y - 1)$ | **M1** | Allow A1√ as FT on expression found for $F(x)$ |
| | *or* $(3/20)y^2 - (1/10)y - 1/20$ | **A2** | Verify $g(y)$ (differentiation may be implied) |
| | $g(y) = G'(y) = (1/10)(3y - 1)$ **AG**<br>[ for $1 \leqslant y \leqslant 3$, $g(y) = 0$ otherwise ] | **M1 A1** | **SC** Missing/incorrect $c$ can earn M1 M1 A1√ M1 (max 4/7) |
| | | **7** | |
| | *OR:* Use of $g(y) = f(x) \times \lvert dx / dy \rvert$ | **(*M1)** | Reference to standard result required (not in syllabus) |
| | $f(x) = (1/20)(3 - 1/y)$ (dep *M1) | **(M1 A1)** | Find $f(x)$ using $x = y^2$ |
| | $dx/dy = 2y$ (dep *M1) | **(M1 A1)** | Find $dx/dy$ using $x = y^2$ |
| | $g(y) = f(x) \times dx/dy = (1/10)(3y - 1)$ **AG**<br>[ for $1 \leqslant y \leqslant 3$, $g(y) = 0$ otherwise ] | **(M1 A1)** | |
| | | **(7)** | |
| 9(ii) | $E(Y) = (1/10)\int(3y^2 - y)\,dy$ | **M1** | Find mean of $Y$ from $\int y\,g(y)\,dy$ |
| | $= (1/10)[y^3 - \frac{1}{2}y^2]_1^3 = 11/5$ *or* 2·2 | **A1** | |
| | | **2** | |

**Q16.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(i) | $F(x) = \int f(x)\,dx = (1/80)\,(2x^{3/2} - 16x^{1/2})$ [+ $c$] | **M1** | Find or state distribution function $F(x)$ for $4 \leqslant x \leqslant 16$<br>using $F(4) = 0$ or $F(16) = 1$ to find $c$ if necessary |
| | $= (1/80)\,(2x^{3/2} - 16x^{1/2} + 16)$<br>*or* $(1/40)\,(x^{3/2} - 8x^{1/2} + 8)$<br>*or* $x^{3/2}/40 - x^{1/2}/5 + 1/5$ (AEF) | **A1** | State $F(x)$ for other values of $x$ |
| | $F(x) = 0$ $(x < 4)$, $F(x) = 1$ $(x > 16)$ | **A1** | |
| | | **3** | |
| 6(ii) | *EITHER:* $G(y)$ [$= P(Y < y) = P(\sqrt{X} < y) = P(X < y^2)$ ]<br>$= F(y^2) = (1/40)\,(y^3 - 8y + 8)$ (AEF) | **M1A1** | Find or state $G(y)$ for $2 \leqslant y \leqslant 4$ from $Y = \sqrt{X}$<br>(allow < or $\leqslant$ throughout; **FT** on constant term) |
| | *OR:* Use $x = y^2$ to find<br>$f(x) = (1/80)\,(3y - 8/y)$ *and* $dx/dy = 2y$ | **(M1A1)** | Find $f(x)$ *and* $dx/dy$ for use in $g(y) = f(x) \times \lvert dx/dy \rvert$ |
| | $g(y)$ [$= G'(y)$] $= (1/40)\,(3y^2 - 8)$ *or* $(3/40)\,y^2 - 1/5$<br>[ for $2 \leqslant y \leqslant 4$, $g(y) = 0$ otherwise ] | **A1** | Find $g(y)$ in simplified form for $2 \leqslant y \leqslant 4$ |
| | | **3** | |

**Q17.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 7(i) | *EITHER:* G($y$) [= P($Y < y$) = P($X^2 < y$)<br>= P($X < y^{1/2}$) = F($y^{1/2}$)] = (1/90) ($y + y^2$) | **M1A1** | Find or state G($y$) for $0 \leqslant x \leqslant 3$ from $Y = X^2$<br>(allow < or $\leqslant$ throughout) |
| | *OR:* Use $x = y^{1/2}$ to find<br>f($x$) = (1/90) ($2x + 4x^3$) = (1/90) ($2y^{1/2} + 4y^{3/2}$)<br>*and* d$x$/d$y$ = 1 / $2y^{1/2}$ | **(M1A1)** | Find f($x$) *and* d$x$/d$y$ for use in g($y$) = f($x$) × \| d$x$/d$y$ \| |
| | g($y$) [= G′($y$)] = (1/90) (1 + 2$y$) | **A1** | Find g($y$) in simplified form |
| | for $0 \leqslant y \leqslant 9$ [g($y$) = 0 otherwise] | **A1** | State corresponding range of $y$ at any stage |
| | | **4** | |
| 7(ii) | E($Y$) = (1/90) $\int (y + 2y^2)\, dy$ | **M1** | Find mean of $Y$ from $\int y\, g(y)\, dy$ |
| | = (1/90) $[\frac{1}{2} y^2 + \frac{2}{3} y^3]_0^9$ = 117/20 *or* 5.85 | **A1** | |
| | | **2** | |

**Q18.** May/June 2017

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 8(i) | $F(x) = \int f(x)\,dx = x^2/8 - x/4$ [+ $c$] | **M1** | Find or state distribution function $F(x)$ for $2 \leqslant x \leqslant 4$ using $F(2) = 0$ or $F(4) = 1$ to find $c$ if necessary |
| | $= x^2/8 - x/4$ *or* $\{(x-1)^2 - 1\}/8$ (AEF) | **A1** | State $F(x)$ for other values of $x$ |
| | $F(x) = 0$ $(x < 2)$, $F(x) = 1$ $(x > 4)$ | **A1** | |
| | **Total:** | **3** | |
| 8(ii) | *EITHER:*<br>$G(y) = P(Y < y) = P((X-1)^3 < y)$<br>$= P(X < 1 + y^{1/3}) = F(1 + y^{1/3})$<br>$= (1 + y^{1/3})^2/8 - (1 + y^{1/3})/4$ *or* $(y^{2/3} - 1)/8$ | **(M1 A1)** | Find or state $G(y)$ for $2 \leqslant x \leqslant 4$ from $Y = (X-1)^3$ (allow < or $\leqslant$ throughout) |
| | *OR:*<br>Use $x = 1 + y^{1/3}$ to find<br>$f(x) = \frac{1}{4} y^{1/3}$ *and* $dx/dy = \frac{1}{3} y^{-2/3}$ | **(M1 A1)** | Find $f(x)$ *and* $dx/dy$ for use in $g(y) = f(x) \times \lvert dx/dy \rvert$ |
| | $g(y)$ [$= G'(y)$] $= (1/12)\, y^{-1/3}$ *or* $1 / (12\, y^{1/3})$ | **A1** | Find $g(y)$ in simplified form |
| | for $1 \leqslant y \leqslant 27$ [$g(y) = 0$ otherwise] | **A1** | State corresponding range of $y$ for $G(y)$ *or* $g(y)$ |
| | **Total:** | **4** | |
| 8(iii) | $(m^{2/3} - 1)/8 = \frac{1}{2}$ | **M1** | Find median value $m$ of $Y$ from $G(m) = \frac{1}{2}$ |
| | $m^{2/3} = 5$, $m = \sqrt{125}$ *or* $5\sqrt{5}$ *or* 11·2 | **M1 A1** | |
| | **Total:** | **3** | |

**Q19.**

| | | | |
|---|---|---|---|
| 9(i) | $(a / \ln 2)\,[-\mathrm{e}^{-x \ln 2}]_0^{\infty} = a / \ln 2$ so $a = \ln 2$ *or* 0·693 | **M1 A1** | State $a$ or find $a$ by equating $\int_0^{\infty} \mathrm{f}(x)\,\mathrm{d}x$ to 1 |
| | **Total:** | **2** | |
| 9(ii) | $\mathrm{E}(X) = 1 / \ln 2$ *or* 1·44 | **B1** | State or find $\mathrm{E}(X)$ |
| | **Total:** | **1** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 9(iii) | $Q$) $= 1 - \mathrm{e}^{-Q \ln 2} =$ ¼ *or* ¾ | **M1** | Formulate equation for either quartile value $Q$ |
| | $Q_1 = (\ln 4/3) / (\ln 2)$ [= 0·415 ] (AEF) | **A1** | Find one [lower] quartile $Q_1$ |
| | $Q_3 = (\ln 4) / (\ln 2)$ [= 2] (AEF) | **A1** | Find other [upper] quartile $Q_3$ |
| | $Q_3 - Q_1$ [= (ln 3) / (ln 2)] = 1·58 [*or* 1·59] | **A1** | Find interquartile range (allow $Q_1 - Q_3$) |
| | **Total:** | **4** | |
| 9(iv) | *EITHER:*<br>$\mathrm{G}(y) = \mathrm{P}(Y < y) = \mathrm{P}(2^X < y)$<br>$= \mathrm{P}(X < (\ln y) / (\ln 2))$<br>$= \mathrm{F}((\ln y) / (\ln 2))$ *or* $\mathrm{F}(\log_2 y)$ (AEF | **(M1 A1** | Find or state $\mathrm{G}(y)$ for $x \geqslant 0$ from $Y = 2^X$<br>(allow < or ≤ throughout) |
| | $= 1 - \mathrm{e}^{-\ln y}$ *or* $1 - 1/y$ | **A1)** | |
| | *OR:*<br>Use $x = (\ln y) / (\ln 2)$ to find *both* | **(M1** | Find $\mathrm{f}(x)$ *and* $\mathrm{d}x/\mathrm{d}y$ for use in $\mathrm{g}(y) = \mathrm{f}(x) \times \lvert \mathrm{d}x/\mathrm{d}y \rvert$ |
| | $\mathrm{f}(x) = (\ln 2)\,\mathrm{e}^{-x \ln 2} = (\ln 2)\,\mathrm{e}^{-\ln y} = (1/y) \ln 2$ | **A1** | |
| | *and* $\mathrm{d}x/\mathrm{d}y = 1 / (y \ln 2)$ | **A1)** | |
| | $\mathrm{g}(y)$ [= $\mathrm{G}'(y)$] $= 1/y^2$ | **A1** | Find $\mathrm{g}(y)$ in simplest form |
| | for $y \geqslant 1$ [$\mathrm{g}(y) = 0$ otherwise] | **A1** | State corresponding range of $y$ for $\mathrm{G}(y)$ or $\mathrm{g}(y)$ |
| | **Total:** | **5** | |

**Q20.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 7(i) | $F(x) = \int f(x)\,dx = -e^{-0.2x} + c = 1 - e^{-0.2x}$ $(x \geq 0)$ | **M1** | State, or integrate and use $F(0) = 0$ or $F(x) \to 1$ as $x \to \infty$ |
| | *and* $F(x) = 0$ ($x < 0$ *or* otherwise) | **A1** | to find, $F(x)$ (**A0** if case $x < 0$ omitted) |
| | | **2** | |
| 7(ii) | $P(X > 2) = 1 - F(2) = e^{-0.4} = 0{\cdot}670$ | **M1 A1** | Find $P(X > 2)$: (**M0** for $F(2)$) |
| | | **3** | |
| 7(iii) | $1 - e^{-0.2m} = ½$ , $e^{0.2m} = 2$ | **M1** | Find median value $m$ from $F(m)$ *or* $1 - F(m) = ½$ |
| | $m = 5 \ln 2$ *or* $3{\cdot}47$ | **M1 A1** | |
| | | **3** | |

**Q21.**

| Question | Mark Scheme Details | | Part Mark | Total |
|---|---|---|---|---|
| **8 (i)** | Find or state distribution function $F(x)$ for $x \geqslant 0$: | $F(x) = \int f(x)\,dx = -e^{-2x} + c$ | **M1** | |
| | Use $F(0) = 0$ to find $F(x)$: | $x) = 0$ $(x < 0)$, $1 - e^{-2x}$ $(x \geqslant 0)$ | **A1** | [2] |
| **(ii)** | Find median value $m$ from $F(m) = ½$: | $1 - e^{-2m} = ½$ , $e^{2m} = 2$ | **M1** | |
| | | $m = ½ \ln 2$ *or* $0{\cdot}347$ | **M1 A1** | [3] |
| **(iii)** | Find or state $G(y)$ from $Y = e^X$ for $x \geqslant 0$: (allow < or ⩽ throughout) | $G(y) = P(Y < y) = P(e^X < y)$ | | |
| | | $= P(X < \ln y) = F(\ln y)$ | | |
| | | $= 1 - e^{-2 \ln y}$ $[= 1 - 1/y^2]$ | **M1 A1** | |
| | Find $g(y)$ by differentiation: | $g(y) = 2 / y^3$ | **A1** | |
| | State corresponding range of $y$: | for $y \geqslant 1$ [$g(y) = 0$ for $y < 1$] | **A1** | [4] |

**Q22.**

| | | | | |
|---|---|---|---|---|
| **7 (i)** | Find F($t$) by integration: | $F(t) = 1 - e^{-\lambda t}$ | **B1** | |
| | Estimate $\lambda$ by equating 1 – F(12) to 280/1000: | $e^{-12\lambda} = 0{\cdot}28$ | | |
| | | $\lambda = (-\ln 0{\cdot}28) / 12 = 0{\cdot}106$ | ***M1 A** | [3] |
| **(ii)** | Estimate mean $\mu$ of $T$: | $\mu = 1/\lambda = 9{\cdot}43$ | **B1** | [1] |
| **(iii)** | Relate new exponential parameter $\lambda'$ to $\lambda$: | $\lambda' = 1/1{\cdot}25\mu = \lambda/1{\cdot}25$ *or* $0{\cdot}8\lambda$ | | |
| | (B1 dependent on *M1) | [= 0·08486 *or* $11{\cdot}78^{-1}$] | **B1** | |
| | Find P($t \leq 12$): | $P(t \leq 12) = 1 - e^{-12\lambda'}$ | **M1** | |
| | | $= 1 - e^{-1{\cdot}018}$ *or* $1 - 0{\cdot}28^{0{\cdot}8}$ | **A1** | |
| | Find percentage: | = 63·9% | **A1** | [4] |

**Q23.**

| | | | | |
|---|---|---|---|---|
| **7 (i)** | Find or state distribution function F($x$) for $2 \leqslant x \leqslant 4$: | $F(x) = \int f(x)\,dx = x^2/12 + c$ | | |
| | Use F(2) = 0 or F(4) = 1 to find F($x$): | $F(x) = x^2/12 - \frac{1}{3}$ [$(2 \leqslant x \leqslant 4)$], 0 (x < 2), 1 (x > 4) | **M1 A1** **A1** | [3] |
| **(ii)** | Find or state G($y$) from $Y = X^3$ for $2 \leqslant x \leqslant 4$: (allow < or $\leqslant$ throughout) (A0 if G($y$) incorrect) | $G(y) = P(Y < y) = P(X^3 < y)$ $= P(X < y^{1/3}) = F(y^{1/3})$ $= y^{2/3}/12 - \frac{1}{3}$ | | |
| | Find g($y$) by differentiation: | $g(y) = (1/18)\, y^{-1/3}$ [for $8 \leqslant y \leqslant 64$, 0 otherwise] | **M1 A1** | [2] |
| **(iii)** | Formulate condition for $k$: (M0 for 7/12 = G($k$)) | (AEF) $7/12 = 1 - G(k) = 1 - k^{2/3}/12 + \frac{1}{3}$ | **M1 A1** **A1** | |
| | Find $k$: | $k^{2/3} = 9$, $k = 27$ | | [3] |

**Q24.**

| | | | | | |
|---|---|---|---|---|---|
| **9** | Find $a$ by equating $\int_2^\infty f(x)dx$ to 1: | $\left[-ae^{-(x-2)}\right]_2^\infty = a = 1$ **AG** | B1 | 1 | |
| | Find or state distribution function F($x$) for $x \geqslant 2$: | $F(x) = \int f(x)\,dx = -ae^{-(x-2)} + c$<br>$F(2) = 0$ so $F(x) = 1 - e^{-(x-2)}$ | M1 A1 | | |
| | State F($x$) for $x < 2$: | $F(x) = 0$ for $x < 2$ | A1 | | |
| | Find median value $m$ from F($m$) = ½: | $1 - e^{-(m-2)} = ½$, $e^{m-2} = 2$<br>$m = 2 + \ln 2$ *or* $2 - \ln ½$ *or* 2.69 | M1 A1 | 5 | |
| **(i)** | Find or state G($y$) from $Y = e^X$ for $x \geqslant 2$: | $G(y) = P(Y < y) = P(e^X < y)$<br>$= P(X < \ln y) = F(\ln y)$<br>$= 1 - e^{-(\ln y - 2)} = 1 - e^2/y$ | M1 A1 | | |
| | Find g($y$) by differentiation (allow $e^2 = 7{\cdot}39$):<br>State corresponding range of $y$: | $g(y) = e^2/y^2$<br>for $y \geqslant e^2$ [$g(y) = 0$ for $y < e^2$] | A1<br>A | 4 | |
| **(ii)** | Find P($Y > 10$):<br>M0 for G(10) or 1 – g(10) | $P(Y > 10) = 1 - G(10)$<br>$= e^2/10$ *or* 0.739 | M1 A1 | 2 | **12** |

**Q25.**

| | | | | | |
|---|---|---|---|---|---|
| **9** | Find $k$ for which $P(X \geqslant k) = 0{\cdot}6$: | $0{\cdot}6 = 1 - F(k)$ | M1 | | |
| | | $= 1 - (k/8 - ¼)$ | M1 | | |
| | | $k = 26/5$ *or* 5·2 | A1 | 3 | |
| | Find G($y$) from $Y = 2 \ln X$ for $2 \leqslant x \leqslant 10$: | $y) = P(Y < y) = P(2 \ln X < y)$ | | | |
| | (allow < or ⩽ throughout) | $= P(X < e^{y/2}) = F(e^{y/2})$ | | | |
| | (result may be stated) | $= e^{y/2}/8 - ¼$ $(2 \ln 2 \leqslant y \leqslant 2 \ln 10)$ | | | |
| | | *or* $(\ln 4 \leqslant y \leqslant \ln 100)$ | | | |
| | | *or* $(1{\cdot}39 \leqslant y \leqslant 4{\cdot}61)$ | M1 A1 | | |
| | State G($y$) for other values of $x$: | 0 $(y < 2 \ln 2)$ *and* 1 $(y > 2 \ln 10)$ | B1 | 3 | |
| | Find g($y$) for $2 \ln 2 \leqslant y \leqslant 2 \ln 10$: | $g(y) = e^{y/2}/16$ | M1 A1 | | |
| | Sketch positive exponential for $2 \ln 2 \leqslant y \leqslant 2 \ln 10$ | | B1 | | |
| | Show $g(y) = 0$ on either side of this interval | | B1 | 4 | **10** |

# 2. Inferential Statistics

**Q.1**

**1** A manager is investigating the times taken by employees to complete a particular task as a result of the introduction of new technology. He claims that the mean time taken to complete the task is reduced by more than 0.4 minutes. He chooses a random sample of 10 employees. The times taken, in minutes, before and after the introduction of the new technology are recorded in the table.

| Employee | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* |
|---|---|---|---|---|---|---|---|---|---|---|
| Time before new technology | 10.2 | 9.8 | 12.4 | 11.6 | 10.8 | 11.2 | 14.6 | 10.6 | 12.3 | 11.0 |
| Time after new technology | 9.6 | 8.5 | 12.4 | 10.9 | 10.2 | 10.6 | 12.8 | 10.8 | 12.5 | 10.6 |

**(a)** Test at the 10% significance level whether the manager's claim is justified. [7]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2022

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** State an assumption that is necessary for this test to be valid. [1]

.....

.....

.....

.....

.....

**Q2.**

**5** Raman is researching the heights of male giraffes in a particular region. Raman assumes that the heights of male giraffes in this region are normally distributed. He takes a random sample of 8 male giraffes from the region and measures the height, in metres, of each giraffe. These heights are as follows.

5.2 5.8 4.9 6.1 5.5 5.9 5.4 5.6

**(a)** Find a 90% confidence interval for the population mean height of male giraffes in this region. [5]

Raman claims that the population mean height of male giraffes in the region is less than 5.9 metres.

**(b)** Test at the 2.5% significance level whether this sample provides sufficient evidence to support Raman's claim. [4]

**Q3.**

**1** The times taken by members of a large quiz club to complete a challenge have a normal distribution with mean $\mu$ minutes. The times, $x$ minutes, are recorded for a random sample of 8 members of the club. The results are summarised as follows, where $\overline{x}$ is the sample mean.

$$\overline{x} = 33.8 \qquad \Sigma(x - \overline{x})^2 = 94.5$$

Find a 95% confidence interval for $\mu$. [4]

**Q4.**

**6** A company has two machines, $A$ and $B$, which independently fill small bottles with a liquid. The volumes of liquid per bottle, in suitable units, filled by machines $A$ and $B$ are denoted by $x$ and $y$ respectively. A scientist at the company takes a random sample of 40 bottles filled by machine $A$ and a random sample of 50 bottles filled by machine $B$. The results are summarised as follows.

$$\Sigma x = 1120 \qquad \Sigma x^2 = 31400 \qquad \Sigma y = 1370 \qquad \Sigma y^2 = 37600$$

The population means of the volumes of liquid in the bottles filled by machines $A$ and $B$ are denoted by $\mu_A$ and $\mu_B$.

**(a)** Test at the 2% significance level whether there is any difference between $\mu_A$ and $\mu_B$. [8]

**(b)** Find the set of values of $\alpha$ for which there would be evidence at the $\alpha$% significance level that $\mu_A - \mu_B$ is greater than 0.25. [4]

**Q5.**

**1** A random sample of 7 observations of a variable $X$ are as follows.

8.26 7.78 7.92 8.04 8.27 7.95 8.34

The population mean of $X$ is $\mu$.

**(a)** Test, at the 10% significance level, the null hypothesis $\mu = 8.22$ against the alternative hypothesis $\mu < 8.22$. [6]

**(b)** State an assumption necessary for the test in part **(a)** to be valid. [1]

**Q6.**

**4** A scientist is investigating the lengths of the leaves of birch trees in different regions. He takes a random sample of 50 leaves from birch trees in region $A$ and a random sample of 60 leaves from birch trees in region $B$. He records their lengths in cm, $x$ and $y$, respectively. His results are summarised as follows.

$$\Sigma x = 282 \qquad \Sigma x^2 = 1596 \qquad \Sigma y = 328 \qquad \Sigma y^2 = 1808$$

The population mean lengths of leaves from birch trees in regions $A$ and $B$ are $\mu_A$ cm and $\mu_B$ cm respectively.

Carry out a test at the 5% significance level to test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

**Q7.**

**1** Farmer *A* grows apples of a certain variety. Each tree produces 14.8 kg of apples, on average, per year. Farmer *B* grows apples of the same variety and claims that his apple trees produce a higher mass of apples per year than Farmer *A*'s trees. The masses of apples from Farmer *B*'s trees may be assumed to be normally distributed.

A random sample of 10 trees from Farmer *B* is chosen. The masses, *x* kg, of apples produced in a year are summarised as follows.

$$\sum x = 152.0 \qquad \sum x^2 = 2313.0$$

Test, at the 5% significance level, whether Farmer *B*'s claim is justified. [6]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

**Q8.**

**3** The heights, $x$ m, of a random sample of 50 adult males from country $A$ were recorded. The heights, $y$ m, of a random sample of 40 adult males from country $B$ were also recorded. The results are summarised as follows.

$$\Sigma x = 89.0 \qquad \Sigma x^2 = 159.4 \qquad \Sigma y = 67.2 \qquad \Sigma y^2 = 113.1$$

Find a 95% confidence interval for the difference between the mean heights of adult males from country $A$ and adult males from country $B$. [8]

**Q9.**

**1** The times taken for students at a college to run 200 m have a normal distribution with mean $\mu$ s. The times, $x$ s, are recorded for a random sample of 10 students from the college. The results are summarised as follows, where $\overline{x}$ is the sample mean.

$$\overline{x} = 25.6 \qquad \Sigma(x-\overline{x})^2 = 78.5$$

**(a)** Find a 90% confidence interval for $\mu$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

A test of the null hypothesis $\mu = k$ is carried out on this sample, using a 10% significance level. The test does not support the alternative hypothesis $\mu < k$.

**(b)** Find the greatest possible value of $k$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q10.**

**4** Manet has developed a new training course to help athletes improve their time taken to run 800 m. Manet claims that his course will decrease an athlete's time by more than 2 s on average. For a random sample of 10 athletes the times taken, in seconds, before and after the course are given in the following table.

| Athlete | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* |
|---|---|---|---|---|---|---|---|---|---|---|
| Before | 150 | 146 | 131 | 135 | 126 | 142 | 130 | 129 | 137 | 134 |
| After | 145 | 138 | 129 | 135 | 122 | 135 | 132 | 128 | 127 | 137 |

Use a *t*-test, at the 5% significance level, to test whether Manet's claim is justified, stating any assumption that you make. [8]

**Q11.**

**1** The number, $x$, of pine trees was counted in each of 40 randomly chosen regions of equal size in country $A$. The number, $y$, of pine trees was counted in each of 60 randomly chosen regions of the same equal size in country $B$. The results are summarised as follows.

$$\Sigma x = 752 \qquad \Sigma x^2 = 14320 \qquad \Sigma y = 1548 \qquad \Sigma y^2 = 40200$$

Find a 95% confidence interval for the difference between the mean number of pine trees in regions of this size in countries $A$ and $B$. [7]

**Q12.**

**6** A scientist is investigating the masses of a particular type of fish found in lakes *A* and *B*. He chooses a random sample of 10 fish of this type from lake *A* and records their masses, *x* kg, as follows.

2.1 1.8 0.9 3.0 2.4 2.6 1.8 2.2 1.9 2.5

The scientist also chooses a random sample of 12 fish of this type from lake *B*, but he only has a summary of their masses, *y* kg, as follows.

$$\Sigma y = 24.48 \qquad \Sigma y^2 = 53.75$$

Test at the 10% significance level whether the mean mass of fish of this type in lake *A* is greater than the mean mass of fish of this type in lake *B*. You should state any assumptions that you need to make for the test to be valid. [10]

© UCLES 2021

**Q13.**

**4** A company has two different machines, $X$ and $Y$, each of which fills empty cups with coffee. The manager is investigating the volumes of coffee, $x$ and $y$, measured in appropriate units, in the cups filled by machines $X$ and $Y$ respectively. She chooses a random sample of 50 cups filled by machine $X$ and a random sample of 40 cups filled by machine $Y$. The volumes are summarised as follows.

$$\Sigma x = 15.2 \quad \Sigma x^2 = 5.1 \quad \Sigma y = 13.4 \quad \Sigma y^2 = 4.8$$

The manager claims that there is no difference between the mean volume of coffee in cups filled by machine $X$ and the mean volume of coffee in cups filled by machine $Y$.

Test the manager's claim at the 10% significance level. [9]

**Q14.**

**5** A large number of children are competing in a throwing competition. The distances, in metres, thrown by a random sample of 8 children are as follows.

19.8 22.1 24.4 21.5 20.8 26.3 23.7 25.0

**(a)** Assuming that distances are normally distributed, test, at the 5% significance level, whether the population mean distance thrown is more than 22.0 metres. [7]

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Find a 95% confidence interval for the population mean distance thrown. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q15.**

**2** A random sample of 40 observations of a random variable $X$ and a random sample of 50 observations of a random variable $Y$ are taken. The resulting values for the sample means, $\overline{x}$ and $\overline{y}$, and the unbiased estimates, $s_x^2$ and $s_y^2$, for the population variances are as follows.

$$\overline{x} = 24.4 \qquad \overline{y} = 17.2 \qquad s_x^2 = 10.2 \qquad s_y^2 = 11.1$$

Find a 90% confidence interval for the difference between the population means of $X$ and $Y$. [5]

**Q16.**

**5** Students at two colleges, $A$ and $B$, are competing in a computer games challenge.

**(a)** The time taken for a randomly chosen student from college $A$ to complete the challenge has a normal distribution with mean $\mu$ minutes. The times taken, $x$ minutes, are recorded for a random sample of 10 students chosen from college $A$. The results are summarised as follows.

$$\Sigma x = 828 \qquad \Sigma x^2 = 68622$$

A test is carried out on the data at the 5% significance level and the result supports the claim that $\mu > k$.

Find the greatest possible value of $k$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** A random sample of 8 students is chosen from college $B$. Their times to complete the same challenge give a sample mean of 79.8 minutes and an unbiased variance estimate of 9.966 minutes$^2$.

Use a 2-sample test at the 5% significance level to test whether the mean time for students at college $B$ to complete the challenge is the same as the mean time for students at college $A$ to complete the challenge. You should assume that the two distributions are normal and have the same population variance. [7]

.....

.....

.....

.....

.....

.....

.....

© UCLES 2020

**Q17.**

**1** Kayla is investigating the lengths of the leaves of a certain type of tree found in two forests *X* and *Y*. She chooses a random sample of 40 leaves of this type from forest *X* and records their lengths, *x* cm. She also records the lengths, *y* cm, for a random sample of 60 leaves of this type from forest *Y*. Her results are summarised as follows.

$$\Sigma x = 242.0 \qquad \Sigma x^2 = 1587.0 \qquad \Sigma y = 373.2 \qquad \Sigma y^2 = 2532.6$$

Find a 90% confidence interval for the difference between the population mean lengths of leaves in forests *X* and *Y*. [7]

**Q18.**

**4** Members of the Sprints athletics club have been taking part in an intense training scheme, aimed at reducing their times taken to run 400 m. For a random sample of 9 athletes from the club, the times taken, in seconds, before and after the training scheme are given in the following table.

| Athlete | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* |
|---|---|---|---|---|---|---|---|---|---|
| Time before | 48.8 | 48.2 | 50.3 | 49.6 | 49.4 | 48.9 | 47.6 | 50.3 | 48.4 |
| Time after | 47.9 | 47.8 | 49.6 | 49.1 | 49.6 | 48.9 | 47.7 | 49.1 | 48.1 |

The organiser of the training scheme claims that on average an athlete's time will be reduced by at least 0.3 seconds.

Test at the 10% significance level whether the organiser's claim is justified, stating any assumption that you make. [8]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2020

**Q19.**

**1** The heights of the members of a large sports club are normally distributed. A random sample of 11 members of the club is chosen and their heights, $x$ cm, are measured. The results are summarised as follows, where $\overline{x}$ denotes the sample mean of $x$.

$$\overline{x} = 176.2 \qquad \Sigma(x-\overline{x})^2 = 313.1$$

Test, at the 5% significance level, the null hypothesis that the population mean height for members of this club is equal to 172.5 cm against the alternative hypothesis that the mean differs from 172.5 cm. [5]

**Q20.**

**6** Nassa is researching the lengths of a particular type of snake in two countries, $A$ and $B$.

**(a)** He takes a random sample of 10 snakes of this type from country $A$ and measures the length, $x$ m, of each snake. He then calculates a 90% confidence interval for the population mean length, $\mu$ m, for snakes of this type, assuming that snake lengths have a normal distribution. This confidence interval is $3.36 \leqslant \mu \leqslant 4.22$.

Find the sample mean and an unbiased estimate for the population variance. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Nassa also measures the lengths, $y$ m, of a random sample of 8 snakes of this type taken from country $B$. His results are summarised as follows.

$$\Sigma y = 27.86 \qquad \Sigma y^2 = 98.02$$

Nassa claims that the mean length of snakes of this type in country $B$ is less than the mean length of snakes of this type in country $A$. Nassa assumes that his sample from country $B$ also comes from a normal distribution, with the same variance as the distribution from country $A$.

Test at the 10% significance level whether there is evidence to support Nassa's claim. [8]

.....

.....

.....

.....

.....

© UCLES 2020

© UCLES 2020

**Q21.**

**8** A large number of runners are attending a summer training camp. A random sample of 6 runners is chosen and their times to run 1500 m at the beginning of the camp and at the end of the camp are recorded. Their times, in minutes, are shown in the following table.

| Runner | *A* | *B* | *C* | *D* | *E* | *F* |
|---|---|---|---|---|---|---|
| Time at beginning of camp | 3.82 | 3.62 | 3.55 | 3.71 | 3.75 | 3.92 |
| Time at end of camp | 3.72 | 3.55 | 3.52 | 3.68 | 3.54 | 3.73 |

The organiser of the training camp claims that a runner's time will improve by more than 0.05 minutes between the beginning and end of the camp. Assuming that differences in time over the two runs are normally distributed, test at the 10% significance level whether the organiser's claim is justified. [8]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2019

**Q22.**

A farmer grows two different types of cherries, Type $A$ and Type $B$. He assumes that the masses of each type are normally distributed. He chooses a random sample of 8 cherries of Type $A$. He finds that the sample mean mass is 15.1 g and that a 95% confidence interval for the population mean mass, $\mu$ g, is $13.5 \leqslant \mu \leqslant 16.7$.

**(i)** Find an unbiased estimate for the population variance of the masses of cherries of Type $A$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

The farmer now chooses a random sample of 6 cherries of Type $B$ and records their masses as follows.

12.2 13.3 16.4 14.0 13.9 15.4

**(ii)** Test at the 5% significance level whether the mean mass of cherries of Type $B$ is less than the mean mass of cherries of Type $A$. You should assume that the population variances for the two types of cherry are equal. [9]

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2019

© UCLES 2019

**Q23.**

**9** A farmer grows large amounts of a certain crop. On average, the yield per plant has been 0.75 kg. The farmer has improved the soil in which the crop grows, and she claims that the yield per plant has increased. A random sample of 10 plants grown in the improved soil is chosen. The yields, $x$ kg, are summarised as follows.

$$\Sigma x = 7.85 \qquad \Sigma x^2 = 6.19$$

**(i)** Test at the 5% significance level whether the farmer's claim is justified, assuming a normal distribution. [7]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** Find a 95% confidence interval for the population mean yield for plants grown in the new soil. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q24.**

A company produces packets of sweets. Two different machines, $A$ and $B$, are used to fill the packets. The manager decides to assess the performance of the two machines. He selects a random sample of 50 packets filled by machine $A$ and a random sample of 60 packets filled by machine $B$. The masses of sweets, $x$ kg, in packets filled by machine $A$ and the masses of sweets, $y$ kg, in packets filled by machine $B$ are summarised as follows.

$$\Sigma x = 22.4 \qquad \Sigma x^2 = 10.1 \qquad \Sigma y = 28.8 \qquad \Sigma y^2 = 16.3$$

A test at the $\alpha\%$ significance level provides evidence that the mean mass of sweets in packets filled by machine $A$ is less than the mean mass of sweets in packets filled by machine $B$. Find the set of possible values of $\alpha$. [12]

© UCLES 2019

© UCLES 2019

## Q25.

**6** A random sample of 9 members is taken from the large number of members of a sports club, and their heights are measured. The heights of all the members of the club are assumed to be normally distributed. A 95% confidence interval for the population mean height, $\mu$ metres, is calculated from the data as $1.65 \leqslant \mu \leqslant 1.85$.

**(i)** Find an unbiased estimate for the population variance. [3]

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** Denoting the height of a member of the club by $x$ metres, find $\Sigma x^2$ for this sample of 9 members. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2019

**Q26.**

**8** A random sample of 8 elephants from region $A$ is taken and their weights, $x$ tonnes, are recorded. (1 tonne = 1000 kg.) The results are summarised as follows.

$$\Sigma x = 32.4 \qquad \Sigma x^2 = 131.82$$

A random sample of 10 elephants from region $B$ is taken. Their weights give a sample mean of 3.78 tonnes and an unbiased variance estimate of 0.1555 tonnes$^2$. The distributions of the weights of elephants in regions $A$ and $B$ are both assumed to be normal with the same population variance. Test at the 10% significance level whether the mean weight of elephants in region $A$ is the same as the mean weight of elephants in region $B$. [9]

**Q27.**

**7** A large number of athletes are taking part in a competition. The masses, in kg, of a random sample of 7 athletes are as follows.

98.1 105.0 92.2 89.8 99.9 95.4 101.2

Assuming that masses are normally distributed, test, at the 10% significance level, whether the mean mass of athletes in this competition is equal to 94 kg. [7]

**Q28.**

**10** The times taken to run 400 metres by students at two large colleges $P$ and $Q$ are being compared. There is no evidence that the population variances are equal. The time taken by a student at college $P$ and the time taken by a student at college $Q$ are denoted by $x$ seconds and $y$ seconds respectively. A random sample of 50 students from college $P$ and a random sample of 60 students from college $Q$ give the following summarised data.

$$\Sigma x = 2620 \qquad \Sigma x^2 = 138\,200 \qquad \Sigma y = 3060 \qquad \Sigma y^2 = 157\,000$$

**(i)** Using a 10% significance level, test whether, on average, students from college $P$ take longer to run 400 metres than students from college $Q$. [9]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2018

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** Find a 90% confidence interval for the difference in the mean times taken to run 400 metres by students from colleges $P$ and $Q$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q29.**

**10** During the summer months, all members of a large swimming club take part in intensive training. The times taken to swim 50 metres at the beginning of the summer and at the end of the summer are recorded for each member of the club. The time taken, in seconds, at the beginning of the summer is denoted by $x$ and the time taken at the end of the summer is denoted by $y$. For a random sample of 9 members the results are shown in the following table.

| Member | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* |
|---|---|---|---|---|---|---|---|---|---|
| $x$ | 38.5 | 40.2 | 32.3 | 35.1 | 36.2 | 41.4 | 32.0 | 38.2 | 38.2 |
| $y$ | 37.4 | 38.1 | 31.6 | 34.7 | 34.2 | 38.6 | 31.8 | 36.3 | 36.8 |

The swimming coach believes that, on average, the time taken by a swimmer to swim 50 metres will decrease by more than one second as a result of the intensive training.

**(i)** Stating suitable hypotheses and assuming a normal distribution, test the coach's belief at the 10% significance level. [8]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** Find a 95% confidence interval for the population mean time taken to swim 50 metres after the intensive training, assuming a normal distribution. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q30.**

**8** The weekly salaries of employees at two large electronics companies, $A$ and $B$, are being compared. The weekly salaries of an employee from company $A$ and an employee from company $B$ are denoted by $\$x$ and $\$y$ respectively. A random sample of 50 employees from company $A$ and a random sample of 40 employees from company $B$ give the following summarised data.

$$\Sigma x = 5120 \qquad \Sigma x^2 = 531\,000 \qquad \Sigma y = 3760 \qquad \Sigma y^2 = 375\,135$$

**(i)** The population mean salaries of employees from companies $A$ and $B$ are denoted by $\$\mu_A$ and $\$\mu_B$ respectively. Using a 5% significance level, test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** State, with a reason, whether any assumptions about the distributions of employees' salaries are needed for the test in part **(i)**. [1]

.....

.....

.....

.....

.....

.....

**Q31.**

**9** There are a large number of students at a particular college. The heights, in metres, of a random sample of 8 students are as follows.

1.75 1.72 1.62 1.70 1.82 1.75 1.68 1.84

You may assume that heights of students are normally distributed.

**(i)** Test, at the 5% significance level, whether the population mean height of students at this college is greater than 1.70 metres. [7]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(ii)** Find a 95% confidence interval for the population mean height of students at this college. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q32.**

**6** The heights, in metres, of a random sample of 8 trees of a particular type are as follows.

14.2 11.3 10.8 8.4 12.8 11.5 12.1 9.2

Assuming that heights of trees of this type are normally distributed, calculate a 95% confidence interval for the mean height of trees of this type. [6]

**Q33.**

**11** In a particular country, large numbers of ducks live on lakes $A$ and $B$. The mass, in kg, of a duck on lake $A$ is denoted by $x$ and the mass, in kg, of a duck on lake $B$ is denoted by $y$. A random sample of 8 ducks is taken from lake $A$ and a random sample of 10 ducks is taken from lake $B$. Their masses are summarised as follows.

$$\Sigma x = 10.56 \qquad \Sigma x^2 = 14.1775 \qquad \Sigma y = 12.39 \qquad \Sigma y^2 = 15.894$$

A scientist claims that ducks on lake $A$ are heavier on average than ducks on lake $B$.

**(i)** Test, at the 10% significance level, whether the scientist's claim is justified. You should assume that both distributions are normal and that their variances are equal. [9]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

A second random sample of 8 ducks is taken from lake $A$ and their masses are summarised as

$$\Sigma x = 10.24 \qquad \text{and} \qquad \Sigma(x - \bar{x})^2 = 0.294,$$

where $\bar{x}$ is the sample mean. The scientist now claims that the population mean mass of ducks on lake $A$ is greater than $p$ kg. A test of this claim is carried out at the 10% significance level, using only this second sample from lake $A$. This test supports the scientist's claim.

**(ii)** Find the greatest possible value of $p$. [5]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q34.**

**7** A farmer grows a particular type of fruit tree. On average, the mass of fruit produced per tree has been 6.2 kg. He has developed a new kind of soil and claims that the mean mass of fruit produced per tree when growing in this new soil has increased. A random sample of 10 trees grown in the new soil is chosen. The masses, $x$ kg, of fruit produced are summarised as follows.

$$\Sigma x = 72.0 \qquad \Sigma x^2 = 542.0$$

Test at the 5% significance level whether the farmer's claim is justified, assuming a normal distribution. [7]

**Q35.**

**9** Two fish farmers $X$ and $Y$ produce a particular type of fish. Farmer $X$ chooses a random sample of 8 of his fish and records the masses, $x$ kg, as follows.

1.2 1.4 0.8 2.1 1.8 2.6 1.5 2.0

Farmer $Y$ chooses a random sample of 10 of his fish and summarises the masses, $y$ kg, as follows.

$$\Sigma y = 20.2 \qquad \Sigma y^2 = 44.6$$

You should assume that both distributions are normal with equal variances. Test at the 10% significance level whether the mean mass of fish produced by farmer $X$ differs from the mean mass of fish produced by farmer $Y$. [10]

**Q36.**

**6** The independent variables $X$ and $Y$ have distributions with the same variance $\sigma^2$. Random samples of $N$ observations of $X$ and $2N$ observations of $Y$ are taken, and the results are summarised by

$$\Sigma x = 4, \quad \Sigma x^2 = 10, \quad \Sigma y = 8, \quad \Sigma y^2 = 102.$$

These data give a pooled estimate of 10 for $\sigma^2$. Find $N$. [5]

**Q37.**

**8** The number, $x$, of beech trees was counted in each of 50 randomly chosen regions of equal size in beech forests in country $A$. The number, $y$, of beech trees was counted in each of 40 randomly chosen regions of the same equal size in beech forests in country $B$. The results are summarised as follows.

$$\Sigma x = 1416 \qquad \Sigma x^2 = 41\,100 \qquad \Sigma y = 888 \qquad \Sigma y^2 = 20\,140$$

Find a 95% confidence interval for the difference between the mean number of beech trees in regions of this size in country $A$ and in country $B$. [9]

**Q38.**

**10** A factory produces bottles of an energy juice. Two different machines are used to fill empty bottles with the juice. The manager chooses a random sample of 50 bottles filled by machine $X$ and a random sample of 60 bottles filled by machine $Y$. The volumes of juice, $x$ and $y$ respectively, measured in appropriate units, are summarised by

$$\Sigma x = 45.5, \quad \Sigma(x - \bar{x})^2 = 19.56, \quad \Sigma y = 72.3, \quad \Sigma(y - \bar{y})^2 = 30.25,$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the volume of juice in the bottles filled by $X$ and $Y$ respectively.

**(i)** Find a 90% confidence interval for the difference between the mean volume of juice in bottles filled by machine $X$ and the mean volume of juice in bottles filled by machine $Y$. [7]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2017

A test at the $\alpha\%$ significance level does not provide evidence that there is any difference in the means of the volume of juice in bottles filled by machine $X$ and the volume of juice in bottles filled by machine $Y$.

**(ii)** Find the set of possible values of $\alpha$. [6]

**Q39.**

**7** A random sample of 9 observations of a normal variable $X$ is taken. The results are summarised as follows.

$$\Sigma x = 24.6 \qquad \Sigma x^2 = 68.5$$

Test, at the 5% significance level, whether the population mean is greater than 2.5. [8]

© UCLES 2016

**[Turn over**

**Q40.**

Petra is studying a particular species of bird. She takes a random sample of 12 birds from nature reserve $A$ and measures the wing span, $x$ cm, for each bird. She then calculates a 95% confidence interval for the population mean wing span, $\mu$ cm, for birds of this species, assuming that wing spans are normally distributed. Later, she is not able to find the summary of the results for the sample, but she knows that the 95% confidence interval is $25.17 \leqslant \mu \leqslant 26.83$. Find the values of $\Sigma x$ and $\Sigma x^2$ for this sample. [6]

Petra also measures the wing spans of a random sample of 7 birds from nature reserve $B$. Their wing spans, $y$ cm, are as follows.

23.2 22.4 27.6 25.3 28.4 26.5 23.6

She believes that the mean wing span of birds found in nature reserve $A$ is greater than the mean wing span of birds found in nature reserve $B$. Assuming that this second sample also comes from a normal distribution, with variance the same as the first distribution, test, at the 10% significance level, whether there is evidence to support Petra's belief. [8]

© UCLES 2016

**Q41.** **8** The coach of a national athletics team carried out an investigation into the effect of high altitude training on the times of 400-metre runners. The times, in seconds, were recorded before and after a six-week period of high altitude training, for a random sample of 8 athletes. The results are given in the following table.

| Athlete | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
|---|---|---|---|---|---|---|---|---|
| Before | 52.3 | 56.2 | 54.3 | 49.2 | 48.4 | 50.1 | 46.8 | 51.1 |
| After | 51.6 | 53.2 | 54.9 | 49.0 | 48.1 | 49.2 | 46.8 | 49.6 |

Stating any assumption that you make, test, at the 2.5% significance level, whether there is an improvement in athletes' times after high altitude training. [8]

© UCLES 2016

**[Turn over**

**Q42.**

**11** The annual salaries of workers at two factories, $A$ and $B$, are to be compared. The salaries, in tens of thousands of dollars, at $A$ and $B$, are denoted by $x$ and $y$ respectively. For a random sample of 40 workers in factory $A$ and a random sample of 50 workers in factory $B$ the results are as follows.

$$\Sigma x = 256.0 \qquad \Sigma x^2 = 1910.8 \qquad \Sigma y = 382.9 \qquad \Sigma y^2 = 3148.8$$

The population mean salaries for $A$ and $B$ are denoted by $\mu_A$ and $\mu_B$ respectively. The population variances for salaries at $A$ and $B$ cannot be assumed to be equal. Test, at the 1% significance level, whether $\mu_B$ is greater than $\mu_A$. [10]

The width of an $\alpha$% confidence interval for $\mu_B - \mu_A$ is found to be 1.82. Find the value of $\alpha$. [4]

© UCLES 2016

**Q43.**

**6** A random sample of 8 observations of a normal random variable $X$ has mean $\bar{x}$, where

$$\bar{x} = 6.246 \quad \text{and} \quad \Sigma(x - \bar{x})^2 = 0.784.$$

Test, at the 5% significance level, whether the population mean of $X$ is less than 6.44. [7]

© UCLES 2016

**Q44.**

**8** The amounts spent on the weekly food shopping by families in the big city $P$ and the small town $Q$ are to be compared. The amounts spent, in dollars, in $P$ and $Q$ are denoted by $x$ and $y$ respectively. For a random sample of 60 families in $P$ and a random sample of 50 families in $Q$, the amounts are summarised as follows.

$$\Sigma x = 9600 \qquad \Sigma x^2 = 1\,560\,000 \qquad \Sigma y = 7200 \qquad \Sigma y^2 = 1\,052\,500$$

Assuming a common population variance, find

**(i)** a pooled estimate for the population variance, [4]

**(ii)** a 95% confidence interval for the difference in the population means in $P$ and $Q$. [5]

© UCLES 2016

**Q45.**

**6** The independent random variables $X$ and $Y$ have distributions with the same variance $\sigma^2$. Random samples of $N$ observations of $X$ and 10 observations of $Y$ are taken, and the results are summarised by

$$\Sigma x = 5, \qquad \Sigma x^2 = 11, \qquad \Sigma y = 10, \qquad \Sigma y^2 = 160.$$

These data give a pooled estimate of 12 for $\sigma^2$. Find $N$. [4]

© UCLES 2015

**[Turn over**

**Q46.**

**7** A random sample of 8 sunflower plants is taken from the large number grown by a gardener, and the heights of the plants are measured. A 95% confidence interval for the population mean, $\mu$ metres, is calculated from the sample data as $1.17 < \mu < 2.03$. Given that the height of a sunflower plant is denoted by $x$ metres, find the values of $\Sigma x$ and $\Sigma x^2$ for this sample of 8 plants. [7]

© UCLES 2015

**[Turn over**

**Q47.**

**10** Young children at a primary school are learning to throw a ball as far as they can. The distance thrown at the beginning of the school year and the distance thrown at the end of the same school year are recorded for each child. The distance thrown, in metres, at the beginning of the year is denoted by $x$; the distance thrown, in metres, at the end of the year is denoted by $y$. For a random sample of 10 children, the results are shown in the following table.

| Child | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* |
|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 5.2 | 4.1 | 3.7 | 5.4 | 7.6 | 6.1 | 3.2 | 4.0 | 3.5 | 8.0 |
| $y$ | 6.2 | 4.8 | 5.0 | 5.6 | 7.7 | 7.0 | 4.0 | 4.5 | 3.6 | 8.5 |

[$\Sigma x = 50.8$, $\Sigma x^2 = 284.16$, $\Sigma y = 56.9$, $\Sigma y^2 = 347.59$, $\Sigma xy = 313.28$.]

A particular child threw the ball a distance of 7.0 metres at the beginning of the year, but he could not throw at the end of the year because he had broken his arm. By finding the equation of an appropriate regression line, estimate the distance this child would have thrown at the end of the year. [5]

The teacher suspects that, on average, the distance thrown by a child increases between the two throws by more than 0.4 metres. Stating suitable hypotheses and assuming a normal distribution, test the teacher's suspicion at the 5% significance level. [8]

© UCLES 2015

**Q48.**

**8** A large number of long jumpers are competing in a national long jump competition. The distances, in metres, jumped by a random sample of 7 competitors are as follows.

6.25 7.01 5.74 6.89 7.24 5.64 6.52

Assuming that distances are normally distributed, test, at the 5% significance level, whether the mean distance jumped by long jumpers in this competition is greater than 6.2 metres. [7]

The distances jumped by another random sample of 8 long jumpers in this competition are recorded. Using the data from this sample of 8 long jumpers, a 95% confidence interval for the population mean, $\mu$ metres, is calculated as $5.89 < \mu < 6.75$. Find the unbiased estimates for the population mean and population variance used in this calculation. [5]

© UCLES 2015

**Q49.**

**10** The times taken, in hours, by cyclists from two different clubs, $A$ and $B$, to complete a 50 km time trial are being compared. The times taken by a cyclist from club $A$ and by a cyclist from club $B$ are denoted by $t_A$ and $t_B$ respectively. A random sample of 50 cyclists from $A$ and a random sample of 60 cyclists from $B$ give the following summarised data.

$$\Sigma t_A = 102.0 \qquad \Sigma t_A^2 = 215.18 \qquad \Sigma t_B = 129.0 \qquad \Sigma t_B^2 = 282.3$$

Using a 5% significance level, test whether, on average, cyclists from club $A$ take less time to complete the time trial than cyclists from club $B$. [10]

A test at the $\alpha$% significance level shows that there is evidence that the population mean time for cyclists from club $B$ exceeds the population mean time for cyclists from club $A$ by more than 0.05 hours. Find the set of possible values of $\alpha$. [4]

© UCLES 2015

**Q50.**

8 Weekly expenses claimed by employees at two different branches, $A$ and $B$, of a large company are being compared. Expenses claimed by an employee at branch $A$ and by an employee at branch $B$ are denoted by \$$x$ and \$$y$ respectively. A random sample of 60 employees from branch $A$ and a random sample of 50 employees from branch $B$ give the following summarised data.

$$\Sigma x = 6060 \qquad \Sigma x^2 = 626\,220 \qquad \Sigma y = 4750 \qquad \Sigma y^2 = 464\,500$$

Using a 2% significance level, test whether, on average, employees from branch $A$ claim the same as employees from branch $B$. [9]

© UCLES 2014

**[Turn over**

**Q51.**

11 The time taken for a randomly chosen student at College $P$ to complete a particular puzzle has a normal distribution with mean $\mu$ minutes. The times, $x$ minutes, are recorded for a random sample of 8 students chosen from the college. The results are summarised as follows.

$$\Sigma x = 42.8 \qquad \Sigma x^2 = 236.0$$

Find a 95% confidence interval for $\mu$. [4]

A test is carried out on this sample data, at the 10% significance level. The test supports the claim that $\mu > k$. Find the greatest possible value of $k$. [3]

A random sample, of size 12, is taken from the students at College $Q$. Their times to complete the puzzle give a sample mean of 4.60 minutes and an unbiased variance estimate of 1.962 minutes$^2$. Use a 2-sample test at the 10% significance level to test whether the mean time for students at College $Q$ to complete the puzzle is less than the mean time for students at College $P$ to complete the puzzle. You should state any assumptions necessary for the test to be valid. [7]

© UCLES 2014

**Q52.**

**6** Employees at a particular company have been working seven hours each day, from 9 am to 4 pm. To try to reduce absence, the company decides to introduce 'flexi-time' and allow employees to work their seven hours each day at any time between 7 am and 9 pm. For a random sample of 10 employees, the numbers of hours of absence in the year before and the year after the introduction of flexi-time are given in the following table.

| Employee | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* |
|---|---|---|---|---|---|---|---|---|---|---|
| Before | 42 | 35 | 96 | 74 | 20 | 5 | 78 | 45 | 146 | 0 |
| After | 34 | 32 | 100 | 72 | 31 | 2 | 61 | 35 | 140 | 0 |

Use a paired sample *t*-test to test, at the 10% significance level, whether the population mean number of hours of absence has decreased, following the introduction of flexi-time. [7]

**Q53.**

**6** A random sample of 50 observations of a random variable $X$ and a random sample of 60 observations of a random variable $Y$ are taken. The results for the sample means, $\bar{x}$ and $\bar{y}$, and the unbiased estimates for the population variances, $s_x^2$ and $s_y^2$, respectively, are as follows.

$$\bar{x} = 25.4 \qquad \bar{y} = 23.6 \qquad s_x^2 = 23.2 \qquad s_y^2 = 27.8$$

A test, at the $\alpha\%$ significance level, of the null hypothesis that the population means of $X$ and $Y$ are equal against the alternative hypothesis that they are not equal is carried out. Given that the null hypothesis is not rejected, find the set of possible values of $\alpha$. [5]

© UCLES 2014

# MS
# Inferential Statistics

**Q.1**

May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 1(a) | Differences:<br>0.6 1.3 0.0 0.7 0.6 0.6 1.8 –0.2 –0.2 0.4 | **M1** | Attempt at differences, allow one error. |
| | $\sum d = 5.6 \quad \sum d^2 = 6.74$<br>$s^2 = \frac{1}{9}\left(6.74 - \frac{5.6^2}{10}\right) = 0.400(4)$ | **M1** | |
| | $H_0$: $\mu_B - \mu_A = 0.4$ and $H_1$: $\mu_B - \mu_A > 0.4$ | **B1** | Allow use of $\mu_d$ |
| | $t = \frac{0.56 - 0.4}{\sqrt{\frac{0.4004}{10}}}$ (0.800) | **M1** | |
| | $t = 0.7996$ (0.800) | **A1** | |
| | Critical value = $t_{0.90}(9) = 1.383$<br>Compare: '0.7996' < 1.383<br>Accept $H_0$ | **M1** | Compare calculated value with 1.383 and correct FT conclusion. |
| | Insufficient evidence to support manager's claim. | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **7** | |
| 1(b) | Distribution of **population** differences is normal | **B1** | |
| | | **1** | |

**Q2.**

May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5(a) | $\sum x = 44.4, \quad \sum x^2 = 247.48$ | **B1** | |
| | $s^2 = \frac{1}{7}\left(247.48 - \frac{44.4^2}{8}\right) = 0.1514\left(= \frac{53}{350}\right)$ | **M1** | |
| | CI: $\frac{44.4}{8} \pm 1.895\sqrt{\frac{0.1514}{8}}$ | **M1** | Correct formula with a t value. |
| | 1.895 | **B1** | 1.895 used. |
| | $5.55 \pm 0.261 = [5.29, 5.81]$ | **A1** | Either form. |
| | | **5** | |
| 5(b) | $H_0 : \mu = 5.9$ and $H_1 : \mu < 5.9$ | **B1** | |
| | $t = \frac{5.55 - 5.9}{\sqrt{\frac{s^2}{8}}} = -2.544$ | **M1** | |
| | Critical value $t_{0.975}(7) = 2.365$<br>Compare '– 2.544' < – 2.365, reject $H_0$. | **M1** | Compare calculated value with $-2.365$ and correct FT conclusion. |
| | Sufficient evidence to support Raman's claim. | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **4** | |

**Q3.**



| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 1 | $s^2 = \frac{94.5}{7} (= 13.5)$ | **B1** | |
| | CI: $33.8 \pm 2.365\sqrt{\frac{s^2}{8}}$ | **M1** | Correct formula with a t value, must be 33.8. |
| | | **B1** | 2.365 used. |
| | $33.8 \pm 3.072 = [30.7, 36.9]$ | **A1** | Either form. |
| | | **4** | |

**Q4.** May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(a) | $s_x^2 = \frac{1}{39}\left(31400 - \frac{1120^2}{40}\right) = 1.0256$<br>$s_y^2 = \frac{1}{49}\left(37600 - \frac{1370^2}{50}\right) = 1.2653$ | **M1** | $\frac{40}{39}$<br>$\frac{62}{49}$ Both required |
| | $s^2 = \frac{1.0256}{40} + \frac{1.2653}{50}$ | **M1** | $\frac{2434}{47775}$<br>Condone $s^2 = \frac{1.03}{40} + \frac{1.27}{50}$ or $0.05115$. |
| | $s^2 = 0.050947$ | **A1** | May be implied by 2.66 as z value. |
| | $H_0 : \mu_a - \mu_b = 0$ and $H_1 : \mu_a - \mu_b \neq 0$ | **B1** | |
| | $z = \frac{\frac{1120}{40} - \frac{1370}{50}}{s} = \frac{0.6}{\sqrt{0.050947}}$ | **M1** | |
| | $z = 2.66$ | **A1** | |
| | Critical value of $z = 2.326$<br>$2.66 > 2.326$<br>Reject $H_0$ | **M1** | Compare their z value with 2.326 and correct FT conclusion. |
| | Sufficient evidence to support difference in **population** means oe | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **8** | |

May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(b) | $z = \frac{0.6 - 0.25}{s}$ | **M1** | Condone 1.54. |
| | $z = 1.55$ | **A1** | |
| | 0.9392 – 0.9395 or 0.0605 – 0.0608 | **A1** | |
| | $\alpha \geqslant 6.05$ | **A1** | Condone 6.05 – 6.08, accept > |
| | | **4** | |

**Q5.**

May/June 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 1(a) | $\bar{x} = \frac{56.56}{7} = 8.08$ | **B1** | Accept unsimplified. |
| | $s^2 = \frac{1}{6}\left(457.275 - \frac{56.56^2}{7}\right) = 0.045033 \left[= \frac{1351}{30000}\right]$ | **M1** | Accept unsimplified. |
| | $t = \frac{8.08 - 8.22}{\sqrt{\frac{s^2}{7}}} = -1.745$ | **M1 A1** | 1.74 – 1.75 |
| | Tabular value 1.440<br>1.745 > 1.440 , so reject $H_0$ | **M1** | Comparison with 1.440 and correct FT conclusion. |
| | There is insufficient evidence to support the hypothesis that the (population) mean is 8.22 OR sufficient evidence to suggest mean is less than 8.22 OR sufficient evidence to suggest $\mu < 8.22$ OE | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | | **6** | |
| 1(b) | Underlying distribution is normal/ population is normal | **B1** | X is normal distributed. |
| | | **1** | |

**Q6.**

May/June 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4 | $s_x^2 = \frac{1}{49}\left(1596 - \frac{282^2}{50}\right) = 0.11265 \left[= \frac{138}{1225}\right]$ | **M1** | One correct, unsimplified. |
| | $s_y^2 = \frac{1}{59}\left(1808 - \frac{328^2}{60}\right) = 0.25311 \left[= \frac{224}{885}\right]$ | **A1** | Both correct to 3sf. |
| | $s^2 = \frac{0.11265}{50} + \frac{0.25312}{60} = 0.006471\ldots$ | **M1 A1** | May be implied. |
| | $z = \frac{\frac{282}{50} - \frac{328}{60}}{\text{their s}}$ | **M1** | |
| | [z] = 2.155 | **A1** | Tolerance 2.15 – 2.16 |
| | Comparison with tabular value 1.96 OE<br>2.155 > 1.96<br>Reject $H_0$ | **M1** | Comparison with 1.96 OE and conclusion. |
| | Sufficient evidence to reject population means are equal OR sufficient evidence to accept population means are not equal OR sufficient evidence to suggest $\mu_A \neq \mu_B$ OE | **A1** | Correct conclusion following correct work. Level of uncertainty in language is used. |
| | | **8** | |

**Q7.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 1 | $s^2 = \frac{1}{9}\left(2313 - \frac{152^2}{10}\right) = 0.28889$ | **M1** | $\frac{13}{45}$ , accept unsimplified. |
| | $H_0 : \mu = 14.8$ $H_1 : \mu > 14.8$ | **B1** | If $\mu$ not used, 'population mean' required.<br>Must see 14.8, must be = and >. |
| | $\left[t =\right] \frac{\frac{152}{10} - 14.8}{s/\sqrt{10}}$ | **M1** | Using unbiased estimate.<br>p-value is 0.0215. |
| | $\left[t =\right] 2.35$ | **A1** | Rounds to 2.35. |
| | Tabular value = 1.833<br>'2.35' > 1.833<br>Reject $H_0$ | **M1** | Comparison with 1.833 and correct FT conclusion. |
| | Sufficient evidence to accept Farmer B's claim oe | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used.<br>Allow $\pm 1$ difference in third significant figure of their t for this mark. |
| | | **6** | |

**Q8.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3 | $\bar{x} = \frac{89}{50} = 1.78 \quad \bar{y} = \frac{67.2}{40} = 1.68$ | **B1** | Implied by 0.1 in the CI formula. |
| | $\sigma_x^2 = \frac{1}{49}\left(159.4 - \frac{89^2}{50}\right) = 0.02 \quad \left(= \frac{1}{50}\right)$ | **M1** | One variance correct unsimplified. |
| | $\sigma_y^2 = \frac{1}{39}\left(113.1 - \frac{67.2^2}{40}\right) = 0.0052308 \quad \left(= \frac{17}{3250}\right)$ | **A1** | Both correctly calculated. |
| | $\sigma^2 = \frac{0.02}{50} + \frac{0.0052308}{40} = \left(\frac{69}{130000}\right)$ | **M1** | Unsimplified expression must be seen, 40 and 50 in correct places. |
| | $[\sigma^2 =] \ 0.0005308$ | **A1** | May be implied by correct final answer. |
| | CI : $1.78 - 1.68 \pm z \times \sqrt{0.0005308}$ | **M1** | Correct form for CI with a z-value. |
| | with $z = 1.96$ | **A1** | |
| | $0.10 \pm 0.0452$ or $[0.0548, 0.145]$ | **A1** | Allow either form. |
| | | **8** | SC Assuming equal variances award B1M0A0, M1A1 (for use of pooled variance formula, gives 0.0135), M1A1A0. |

**Q9.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 1(a) | $s^2 = \frac{78.5}{9} = 8.7222$ | **B1** | $\frac{157}{18}$ |
| | CI: $25.6 \pm t\sqrt{\frac{s^2}{10}}$ | **M1** | Correct expression with a t value. |
| | With t = 1.833 | **A1** | With correct t value. |
| | $25.6 \pm 1.71(2)$ or [23.9, 27.3] | **A1** | Accept in either form, ISW. Accept inequality form. |
| | | **4** | |
| 1(b) | $\frac{25.6 - k}{\sqrt{\frac{s^2}{10}}} \geqslant t$ | **M1** | |
| | With t = −1.383 | **A1** | Allow 1.383. |
| | 26.9 | **A1** | CAO |
| | | **3** | |

**Q10.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4 | $H_0 : \mu_B - \mu_A = 2$ and $H_1 : \mu_B - \mu_A > 2$ | **B1** | Or use of $\mu_d$. |
| | Differences: 5 8 2 0 4 7 −2 1 10 −3 | **M1** | Allow one error. |
| | $\sum d = 32, \quad \sum d^2 = 272$<br>$\bar{d} = 3.2, \ s^2 = \frac{1}{9}\left(272 - \frac{32^2}{10}\right) = 18.84 \ \left(= \frac{848}{45}\right)$ | **M1** | Sample mean and variance. |
| | $t = \frac{3.2 - 2}{\sqrt{\frac{s^2}{10}}} = 0.874$ | **M1 A1** | |
| | Compare with tabular value 1.833: 0.874 < 1.833.<br>Accept $H_0$. | **M1** | Compare their value with 1.833 and conclusion. |
| | Insufficient evidence to support claim. | **A1** | Correct conclusion, in context, following correct work.<br>Level of uncertainty in language is used. |
| | Assumption: population differences are normally distributed | **B1** | |
| | | **8** | |

**Q11.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 1 | $s_x^2 = \frac{1}{39}\left(14320 - \frac{752^2}{40}\right) = 4.677 = \frac{304}{65}$ | **M1** | One correct unsimplified. |
| | $s_y^2 = \frac{1}{59}\left(40200 - \frac{1548^2}{60}\right) = 4.434 = \frac{1308}{295}$ | **A1** | Both correct to at least 3sf. |
| | $s^2 = \frac{4.6769}{40} + \frac{4.4339}{60}$ | **M1** | |
| | $0.1908 \left(= \frac{3659}{19175}\right)$ | **A1** | |
| | $CI = \left(\frac{752}{40} - \frac{1548}{60}\right) \pm 1.96s$ | **M1A1** | Correct form with a z-value with 1.96 s can be unsimplified. |
| | $-7.0 \pm 0.856$ = $[-7.86, -6.14]$ | **A1** | Accept in either form, ISW.<br>Accept [6.14, 7.86].<br>Accept inequality form. |
| | | **7** | |

**Q12.**



| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6 | $H_0 : \mu_A = \mu_B$ and $H_1 : \mu_A > \mu_B$ | **B1** | |
| | $\sum x = 21.2 \quad \sum x^2 = 47.92$<br>$s_x^2 = \frac{1}{9}\left(47.92 - \frac{21.2^2}{10}\right) = 0.33067 \ = \frac{124}{375}$<br>$s_y^2 = \frac{1}{11}\left(53.75 - \frac{24.48^2}{12}\right) = 0.34644 \ = \frac{9527}{27500}$ | **M1 A1** | |
| | $s^2 = \frac{9 \times 0.33067 + 11 \times 0.34644}{10 + 12 - 2} = 0.3393$ | **M1 A1** | |
| | $t = \frac{2.12 - 2.04}{s\sqrt{\frac{1}{10} + \frac{1}{12}}} = 0.321$ | **M1 A1** | Accept 0.321 – 0.322. |
| | Critical value: 1.325<br>$0.321 < 1.325$ accept $H_0$. | **M1** | Compare their value with 1.325 and correct FT conclusion. |
| | Insufficient evidence that mean of A is greater than mean of B | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | EITHER: Distributions are normal and equal variances<br>OR: Distributions are normal | **B1** | Assumptions consistent with method used.<br>Accept 'population is normal'.<br>'Distribution of difference of means is normal'. |
| | **Alternative method for question 6** | | |
| | $H_0 : \mu_A = \mu_B$ and $H_1 : \mu_A > \mu_B$ | **B1** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| | $\sum x = 21.2 \quad \sum x^2 = 47.92$<br>$s_x^2 = \frac{1}{9}\left(47.92 - \frac{21.2^2}{10}\right) = 0.33067 \ = \frac{124}{375}$<br>$s_y^2 = \frac{1}{11}\left(53.75 - \frac{24.48^2}{12}\right) = 0.34644 \ = \frac{9527}{27500}$ | **M1**<br>**A1** | For one correct unsimplified.<br>For both correct. |
| | $s^2 = \frac{0.33067}{10} + \frac{0.34644}{12} = 0.061936$ | **M1 A1** | |
| | $t = \frac{2.12 - 2.04}{s} = 0.321(5)$ | **M1 A1** | Accept 0.321 – 0.322. |
| | Critical value: 1.325<br>$0.321 < 1.325$ accept $H_0$. | **M1** | Compare their value with 1.325 and correct FT conclusion. |
| | Insufficient evidence that mean of A is greater than mean of B. | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | EITHER: Distributions are normal and equal variances<br>OR: Distributions are normal. | **B1** | Assumptions consistent with method used.<br>Accept 'population is normal'.<br>'Distribution of difference of means is normal'. |
| | | **10** | |

**Q13.** May/June 2020

| Question | Answer | Marks |
|---|---|---|
| 4 | $H_0$: $\mu_x = \mu_y$ $H_1$: $\mu_x \neq \mu_y$ | **B1** |
| | $s_x^2 = \frac{1}{49}\left(5.1 - \frac{15.2^2}{50}\right) = 0.0097796$: $s_y^2 = \frac{1}{39}\left(4.8 - \frac{13.4^2}{40}\right) = 0.007974$ | **M1A1** |
| | $s^2 = \frac{0.00977959}{50} + \frac{0.007974}{40} = 0.0003949$ | **M1A1** |
| | $z = \frac{0.304 - 0.335}{\sqrt{0.0003949}} = (-)1.56$ | **M1A1** |
| | Compare with 1.645 | **M1** |
| | Accept $H_0$: insufficient evidence to reject manager's claim | **A1** |
| | | **9** |

**Q14.**

May/June 2020

| Question | Answer | Marks |
|---|---|---|
| 5(a) | $\sum x = 183.6,\ \sum x^2 = 4249.08,\ \ \bar{x} = 22.95$ | **B1** |
| | $s^2 = \frac{1}{7}\left(4249.08 - \frac{183.6^2}{8}\right) = 5.066$ | **M1** |
| | $H_0$: $\mu = 22.0$, $H_1$: $\mu > 22.0$ | **B1** |
| | $t = \frac{22.95 - 22.0}{\sqrt{\frac{s^2}{8}}} = 1.194$ | **M1A1** |
| | Compare t with correct tabular value 1.895 | **M1** |
| | Accept $H_0$: mean distance thrown is not more than 22.0 m | **A1** |
| | | **7** |
| 5(b) | $22.95 \pm t\sqrt{\frac{s^2}{8}}$ | **M1** |
| | With t = 2.365 | **B1** |
| | [21.1, 24.8] | **A1** |
| | | **3** |

**Q15.**

| Question | Answer | Marks |
|---|---|---|
| 2 | $s^2 = \frac{10.2}{40} + \frac{11.1}{50} = 0.477$ | **M1A1** |
| | CI = $(24.4 - 17.2) \pm zs$ | **M1** |
| | = $(24.4 - 17.2) \pm 1.645\sqrt{0.477}$ | **A1** |
| | = [6.06, 8.34] | **A1** |
| | | **5** |

**Q16.** May/June 2020

| Question | Answer | Marks |
|---|---|---|
| 5(a) | $\bar{x} = \frac{828}{10} = 82.8 \quad s^2 = \frac{1}{9}\left(68622 - \frac{828^2}{10}\right) = 7.0667$ | **B1** |
| | $\frac{82.8 - k}{\sqrt{\frac{s^2}{10}}} \geqslant t$ where $t = 1.833$<br>(**M1** if equality, **A1** for inequality and correct t value) | **M1 A1** |
| | $k \leqslant 81.259 \quad k \leqslant 81.3$ | **A1** |
| | | **4** |
| 5(b) | $H_0$: $\mu_A = \mu_B$ $H_1$: $\mu_A \neq \mu_B$ | **B1** |
| | Pooled variance $= \frac{9 \times 7.0667 + 7 \times 9.966}{10 + 8 - 2} = s_p^2$ | **M1** |
| | $= 8.335$ | **A1** |
| | $t = \frac{82.8 - 79.8}{s_p\sqrt{\frac{1}{10} + \frac{1}{8}}} = 2.19$ | **M1A1** |
| | Compare with 2.12 (t(16, 0.975)) and reject $H_0$ | **M1** |
| | Population means are not the same | **A1** |
| | | **7** |

**Q17.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 1 | $s_x^2 = \frac{1}{39}\left(1587.0 - \frac{242.0^2}{40}\right) = 3.15128 \quad \left(= \frac{1229}{390}\right)$<br>$s_y^2 = \frac{1}{59}\left(2532.6 - \frac{373.2^2}{60}\right) = 3.58129 \quad \left(= \frac{26412}{7375}\right)$ | **M1 A1** | Either unsimplified<br>Both correct |
| | $s^2 = \frac{3.15128}{40} + \frac{3.58129}{60} = 0.13847$ | **M1 A1** | Pooled variance is M0A0<br>or $s = 0.3721$ |
| | $6.05 - 6.22 \pm zs$ | **M1** | FT their s, must be a z value |
| | $= 0.17 \pm 1.645\sqrt{0.13847}$ | **A1** | With 1.645 |
| | $= [-0.442, 0.782]$ or $[-0.782, 0.442]$ | **A1** | |
| | | **7** | |

**Q18.**

| | | | |
|---|---|---|---|
| 4 | Assume (population) differences are normally distributed | **B1** | |
| | $H_0$: $\mu_X - \mu_Y = 0.3$ $H_1$: $\mu_X - \mu_Y > 0.3$ | **B1** | |
| | Diff: 0.9 0.4 0.7 0.5 –0.2 0 –0.1 1.2 0.3 | **M1** | Signed differences |
| | $\sum d = 3.7, \ \sum d^2 = 3.29$<br>$\bar{d} = 0.411, \ s^2 = \frac{1}{8}\left(3.29 - \frac{3.7^2}{9}\right) = 0.2211$ | **M1** | $s = 0.4702$ |
| | $t = \frac{0.411 - 0.3}{\sqrt{\frac{0.2211}{9}}}$<br>$= 0.708$ or 0.709 | **M1 A1** | |
| | '0.708' < 1.397 | **M1** | Compare their 0.708 with 1.397 |
| | Accept $H_0$<br>Insufficient evidence to support claim | **A1 FT** | In context, except possibly hypotheses<br>Level of uncertainty in language used. No contradictions |
| | | **8** | |

**Q19.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 1 | $s^2 = \frac{313.1}{10} = 31.3$ | **B1** | Can be implied |
| | $t = \frac{176.2 - 172.5}{s / \sqrt{11}}$ | **M1** | |
| | = 2.19 | **A1** | |
| | '2.19' < 2.228 | **M1** | Comparison of their t value with 2.228 |
| | accept $H_0$:<br>insufficient evidence to reject mean height is 172.5<br>OR<br>sufficient evidence to accept mean height is 172.5 | **A1 FT** | Correct conclusion in context, level of uncertainty in language used. No contradictions.<br>CWO<br>Do not accept: 'mean height is 172.5'<br>Do not accept use of symbols that are undefined e.g. insufficient evidence to reject $\mu = 172.5$ |
| | | **5** | |

**Q20.** October/November 2020

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(a) | $\bar{x} = \frac{1}{2}(4.22 + 3.36) = 3.79$ | **B1** | |
| | $4.22 - 3.36 = \frac{2ts}{\sqrt{10}}$ | **M1** | Using z value implies M0 |
| | With t = 1.833 | **A1** | |
| | ( s = 0.7418) variance = 0.55(0) | **A1** | |
| | | **4** | |
| 6(b) | $H_0$: $\mu_A = \mu_B$ $H_1$: $\mu_A > \mu_B$ | **B1** | Not $\bar{x}$ |
| | $s^2 = \frac{1}{7}\left(98.02 - \frac{27.86^2}{8}\right) = 0.142(507)$ | **M1** | $= \frac{1}{7}(98.02 - 97.02)$ |
| | Pooled estimate = $\frac{9 \times 0.550316 + 7 \times 0.142507}{16}$ | **M1** | |
| | $= 0.372$ | **A1** | |
| | $t = \frac{\frac{27.86}{8} - 3.79}{\sqrt{0.372}\sqrt{\frac{1}{10} + \frac{1}{8}}} = (-)1.063$ | **M1A1** | |
| | '1.063' < 1.337 oe | **M1** | Compare their value with 1.337 |
| | Accept $H_0$<br>Nassa's claim is not supported | **A1ft** | Correct conclusion in context, level of uncertainty in language used. No contradictions. |
| | | **8** | |

**Q21.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 8 | $H_0$: $\mu_b - \mu_e = 0{\cdot}05$, $H_1$: $\mu_b - \mu_e > 0{\cdot}05$ (AEF) | **B1** | State both hypotheses (B0 for $\bar{x}$ … ) |
| | $d_i$: 0·10 0·07 0·03 0·03 0·21 0·19 (or in sec) | **M1** | Consider differences $d_i$ from e.g. $x_b - y_e$ |
| | $\bar{d} = 0{\cdot}63 / 6 = 0{\cdot}105$ (or 6·3 sec) | **B1** | Find sample mean |
| | $s^2 = (0{\cdot}0969 - 0{\cdot}63^2/6) / 5$<br>[ = 123/20 000 or 0·00615 or $0{\cdot}0784^2$ ](or 22·14) | **M1** | Estimate population variance<br>(allow biased here: [41/8000 or 0·005125 or $0{\cdot}0716^2$ ]) |
| | $t_{5,\,0.9} = 1{\cdot}476$ (to 3 sf) | **B1** | State or use correct tabular t-value |
| | $t = (\bar{d} - 0{\cdot}05) / (s/\sqrt{6}) = 1{\cdot}72$ | **M1 A1** | Find value of t<br>(or compare $\bar{d} - 0{\cdot}05 = 0{\cdot}055$ with $(t_{5,\,0.9})s/\sqrt{6} = 0{\cdot}0473$) |
| | [Reject $H_0$:] Evidence for organiser's belief<br>or times improve by more than 0·05 min (AEF) | **B1** | **FT** on both t-values<br>Consistent conclusion |
| | | | **SC** Wrong type of hypothesis test can earn only<br>B1 for hypotheses<br>B1FT for conclusion (max 2/8) |
| | | **8** | |

**Q22.** May/June 2019

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 11O(i) | $ts_A / \sqrt{8} = ½ (16·7 – 13·5)$ [= 1·6] | **M1** | Relate $s_A$ to semi-width of confidence interval |
| | $t_{7, 0.975} = 2·365$ (to 3 s.f.) | **A1** | State or use correct tabular t value |
| | [$s_A = \sqrt{8} \times 1·6 / 2·365 = 1·9135$], $s_A^2 = 3·66$[16] | **A1** | Hence find unbiased estimate of A's population variance |
| | | **3** | |
| 11O(ii) | $H_0: \mu_A = \mu_B$ , $H_1: \mu_A > \mu_B$ (AEF) | **B1** | State hypotheses (B0 for $\bar{x}$ …) |
| | [$\bar{x}_A = 15.1$], $\bar{x}_B = 85.2 / 6 = 14.2$ | **B1** | Find sample mean for B |
| | $s_B^2 = (1221·06 – 85·2^2/6) / 5$<br>$= 561/250$ or $2·244$ or $1·498^2$ (all to 3 s.f.) | **M1** | Estimate or imply population variance for B<br>(allow biased here: $1·87$ or $1·367^2$) |
| | $s^2 = (7 s_A^2 + 5 s_B^2) / 12 = 3·0709$ or $1·752^2$ | **M1 A1** | Estimate (pooled) common variance ($s_B^2$ not needed explicitly) |
| | $t_{12, 0.95} = 1·782$ | **B1** | State or use correct tabular t value |
| | [-] $t = (\bar{x}_A - \bar{x}_B) / (s \sqrt{(1/8 + 1/6)}) = 0·951$<br>$t < 1·78$ so [accept $H_0$] | **M1 A1** | Find value of t (or can compare $\bar{x}_A - \bar{x}_B = 0·9$ with 1·69)<br>Correct conclusion |
| | mean mass of B not less than mean mass of A (AEF) | **B1** | |
| | | **9** | |
| | | | **SC1:** Implicitly taking $s_A^2$, $s_B^2$ as unequal population variances<br>(may also earn first B1 B1 M1)<br>$z = (\bar{x}_A - \bar{x}_B) / \sqrt{(s_A^2/8 + s_B^2/6)}$<br>$= 0·9 / \sqrt{(0·8317)} = 0·987$<br>$z < 1·645$ so |
| | | | **DepSC1**: mean mass of B not less than mean mass of A (AEF)<br>Comparison with $z_{0.95}$ and conclusion (FT on z)<br>(can earn at most 5/9) |

**Q23.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 9(i) | $\bar{x} = 0{\cdot}785$ | **B1** | Find sample mean |
| | $s^2 = (6{\cdot}19 - 7{\cdot}85^2/10) / 9$<br>[ = 37/12 000 or 0·003083 or $0{\cdot}05553^2$ ] | **M1** | Estimate population variance<br>(allow biased here: 0·002775 or $0{\cdot}05268^2$) |
| | $H_0: \mu = 0{\cdot}75$, $H_1: \mu > 0{\cdot}75$ (AEF) | **B1** | State hypotheses (B0 for $\bar{x}$ …) |
| | $t_{9,\,0.95} = 1{\cdot}83$[3] | **B1** | State or use correct tabular t-value |
| | $t = (\bar{x} - 0{\cdot}75)/(s/\sqrt{10}) = 1{\cdot}99$<br>[Reject $H_0$ and accept $H_1$] | **M1 A1** | Find value of t (or compare $\bar{x}$ with 0·75 + 0·032 = 0·782)<br>Consistent conclusion |
| | Claim (of yield per plant increased) is justified (AEF) | **B1** | **FT** on both t-values (must be t value) |
| | | **7** | |
| 9(ii) | $\bar{x} \pm t\sqrt{(s^2/10)}$ | **M1** | Find confidence interval (must be a t value) |
| | $t_{9,\,0.975} = 2{\cdot}26$[2] | **B1** | State or use correct tabular value of t |
| | $0{\cdot}785 \pm 0{\cdot}04$ or [0·745, 0·825] | **A1** | Evaluate confidence interval (either form) |
| | | **3** | |

**Q24.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 11O | $\bar{x} - \bar{y} = 0{\cdot}448 - 0{\cdot}48 = [-]\ 0{\cdot}032$ | **B1** | Find difference in sample means (either sign; may be implied) |
| | $s_X^2 = (10{\cdot}1 - 22{\cdot}4^2/50) / 49 = 0{\cdot}0648 / 49 \quad [= 0{\cdot}001322]$ | **M1 A1** | Estimate both population variances (may be implied) |
| | $s_Y^2 = (16{\cdot}3 - 28{\cdot}8^2/60) / 59 = 2{\cdot}476 / 59 \quad [= 0{\cdot}04197]$ | **A1** | (allow biased here: 0·001296 and 0·04127) |
| | $s_C^2 = s_X^2/50 + s_Y^2/60$ | **M1** | Estimate combined variance |
| | $= 0{\cdot}0007259$ or $0{\cdot}02694^2$ (to 3 s.f. throughout) | **A1** | |
| | $z = 0{\cdot}032 / s_C$ [or $0{\cdot}032 / s_P\sqrt{(1/50 + 1/60)}$]<br>$= 1{\cdot}188$ [or 1·0895] | **M1 A1** | Find value of z (either sign) |
| | $\Phi(z) = 0{\cdot}8826$ [or 0·8620] | **A1** | Find $\Phi(z)$ |
| | $100 \times (1 - \Phi(z)) = 11{\cdot}7$ [or 13·8] | **M1 A1** | Find limiting value for $\alpha$, based on one-tail test<br>(M0 for basing on two-tail test) |
| | $\alpha >$ (or $\geqslant$) 11·7 [or 13·8] | **A1** | Find set of possible values of $\alpha$ Allow 11.8 if 11.74 seen<br>(Misreading $\alpha$% as $\alpha$ loses only last A1) |

**Q25.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(i) | $t\sqrt{(s^2/9)} = \frac{1}{2}(1{\cdot}85 - 1{\cdot}65)$ [= 0·1] | **M1** | Find estimate $s^2$ of population variance (must be t) |
| | $t_{8,\,0.975} = 2{\cdot}306$ (to 3 s.f.) | **A1** | Use of correct tabular t-value |
| | $s^2 = 9 \times 0{\cdot}04337^2 = 0{\cdot}0169$ or $0{\cdot}130[1]^2$ | **A1** | |
| | | **3** | |
| 6(ii) | $\bar{x} = \frac{1}{2}(1{\cdot}65 + 1{\cdot}85) = 1{\cdot}75$<br>or $\Sigma x = 9 \times 1{\cdot}75 = 15{\cdot}75$ | **M1 A1** | Find sample mean $\bar{x}$ |
| | $s^2 = (\Sigma x^2 - 9 \times \bar{x}^2) / 8$<br>or $\{\Sigma x^2 - (\Sigma x)^2 / 9\} / 8$ | **M1** | or $\Sigma x$ |
| | $\Sigma x^2 = 8 \times 0{\cdot}0169 + 15{\cdot}75^2/9 = 27{\cdot}7$ | **A1** | Find $\Sigma x^2$ from $s^2$ |
| | | **4** | |

**Q26.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 8 | $\bar{x}_A = 32{\cdot}4 / 8 = 4{\cdot}05$ | **B1** | Find sample mean for A |
| | $s_A^2 = (131{\cdot}82 - 32{\cdot}4^2/8) / 7$<br>$s_A^2 = 3/35$ (or 0·08571 or $0{\cdot}2928^2$ both to 3 s.f.) | **M1** | Estimate or imply popln. variance for A |
| | $H_0$: $\mu_A = \mu_B$ , $H_1$: $\mu_A \neq \mu_B$ | **B1** | State hypotheses. AEF |
| | $s^2 = (7 s_A^2 + 9 s_B^2) / 16 = 0{\cdot}12497$ or $0{\cdot}3535^2$ or $\frac{3999}{32000}$ | **M1 A1** | Estimate (pooled) common variance |
| | $t_{16,\,0.95} = 1{\cdot}746$ | **B1*** | State or use correct tabular t value |
| | [–] $t = (x_A - \bar{x}_B) / s\sqrt{(1/8 + 1/10)}$ | **M1** | |
| | $= 0{\cdot}27 / 0{\cdot}1677 = 1{\cdot}61$ | **A1** | Find value of t (or can compare $\bar{x}_A - \bar{x}_B = 0{\cdot}27$ with 0·293) |
| | $t < 1{\cdot}75$ so [accept $H_0$] mean masses are the same | **DB1** | Correct conclusion (FT on t, dep B1*). AEF |
| | | **9** | |

**Q27.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 7 | $\bar{x} = 681{\cdot}6/7 = 97{\cdot}37$ (allow 97·4 for this B1) | **B1** | Find sample mean |
| | $s^2 = (66\,536{\cdot}1 - 681{\cdot}6^2/7) / 6$<br>$[ = 27{\cdot}96$ or $5{\cdot}287^2 ]$ | **M1** | Estimate population variance<br>(allow biased here: 23·96 or $4{\cdot}895^2$) |
| | $H_0$: $\mu = 94$, $H_1$: $\mu \neq 94$ (AEF) | **B1** | State hypotheses (B0 for $\bar{x}$ …) |
| | $t = (\bar{x} - 94)/(s/\sqrt{7}) = 1{\cdot}69$ | **M1 A1** | Find value of t |
| | $t_{7,\,0.95} = 1{\cdot}94[3]$ | **B1** | State or use correct tabular t-value<br>(or can compare $\bar{x}$ with 94 + 3·88 = 97·88 ) |
| | [Accept $H_0$:] Mean mass is equal to 94 kg (AEF) | **B1√** | Consistent conclusion (FT on both t-values) |
| | | **7** | |

**Q28.**

May/June 2018

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 10(i) | $H_0$: $\mu_P = \mu_Q$ , $H_1$: $\mu_P > \mu_Q$ | **B1** | State hypotheses (B0 for $\bar{x}$ …) |
| | $s_P^2 = (138\,200 - 2620^2/50) / 49 = 18{\cdot}61$ or $912/49$ and<br>$s_Q^2 = (157\,000 - 3060^2/60) / 59 = 15{\cdot}93$ or $940/59$ | **M1 A1** | Estimate both popn. variances (to 3 s.f.)<br>(allow biased here: 18·24 and 15·67) |
| | $s^2 = s_P^2/50 + s_Q^2/60 = 0{\cdot}6378$ or $0{\cdot}7986^2$ | **M1 A1** | Estimate combined variance (to 3 s.f.; may be implicit) |
| | $z_{0.9} = 1{\cdot}28[2]$ | ***B1** | State or use correct tabular z value |
| | $z = (\bar{y} - \bar{x}) / s = (52{\cdot}4 - 51) / s = 1{\cdot}75$<br>$z >$ tabular value so [accept $H_1$ and] | **M1 A1** | Calculate value of z (or –z)<br>(or can compare $\bar{y} - \bar{x} = 1{\cdot}02$ with 1·4) |
| | college P students take more time (AEF) | **B1√** | Correctly stated conclusion (FT on z, dep *B1) |
| | [**SC**: $s^2 = (912 + 940)/108 = 17{\cdot}15$ or $4{\cdot}141^2$<br>$z = 1{\cdot}4 / s\sqrt{(1/50 + 1/60)} = 1{\cdot}77$ ] | | **SC:** Using pooled estimate of common variance can earn<br>B1 M1 A1 (may be implied) M0 *B1 M0 B1√ (max 5/9) |
| | | **9** | |
| 10(ii) | $\bar{x} - \bar{y} \pm z\,s$ (or $\bar{y} - \bar{x} \pm z\,s$) | **M1** | Find confidence interval for difference |
| | $z_{0.95} = 1{\cdot}64[5]$ | **A1** | Use appropriate tabular value |
| | $1{\cdot}4 \pm 1{\cdot}31$ or [0·09, 2·71]<br>or $-1{\cdot}4 \pm 1{\cdot}31$ or [–2·71, –0·09]<br>[**SC**: e.g. $\bar{x} - \bar{y} \pm z\,s\sqrt{(1/50 + 1/60)} = 1{\cdot}4 \pm 1{\cdot}30$ ] | **A1** | Evaluate confidence interval (either form)<br><br>**SC**: Allow M1 A1 if common variance used (max 2/3) |
| | | **3** | |

**Q29.** May/June 2018

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| **10(i)** | $H_0$: $\mu_x - \mu_y = 1$, $H_1$: $\mu_x - \mu_y > 1$ (AEF) | **B1** | State both hypotheses (B0 for $\bar{x}$ …) |
| | $d_i$: 1·1 2·1 0·7 0·4 2·0 2·8 0·2 1·9 1·4 | **M1** | Consider differences $d_i$, e.g. $x_i - y_i$ |
| | $\bar{d} = 12{\cdot}6 / 9 = 1{\cdot}4$ | **B1** | Find sample mean |
| | $s^2 = (23{\cdot}72 - 12{\cdot}6^2/9) / 8$<br>$[= 0{\cdot}76$ or $0{\cdot}8718^2]$ | **M1** | Estimate population variance<br>(allow biased here: $[0{\cdot}6756$ or $0{\cdot}8219^2]$) |
| | $t_{8,\,0.9} = 1{\cdot}397$ or $1{\cdot}40$ | **B1** | State or use correct tabular t-value |
| | $t = (\bar{d} - 1) / (s/\sqrt{9}) = 1{\cdot}38$<br>[Accept $H_0$:] No evidence for coach's belief | **M1 A1** | Find t (or compare $\bar{d} - 1 = 0{\cdot}4$ with $t_{8,\,0.9}\, s/\sqrt{9} = 0{\cdot}406$)<br>Consistent conclusion (FT on both t-values) |
| | or time will not decrease by more than 1 s (AEF) | **B1√** | |
| | | | **SC** Wrong (hypothesis) test can earn only<br>B1 for hypotheses<br>B1√ for conclusion (max 2/8) |
| | | **8** | |
| **10(ii)** | $s_y^2 = (11\,395{\cdot}79 - 319{\cdot}5^2/9) / 8$<br>$= 6{\cdot}693$ or $2{\cdot}587^2$ (to 3 sf) | **B1** | Estimate population variance using y's sample<br>(allow use of biased here: $5{\cdot}950$ or $2{\cdot}439^2$) |
| | $319{\cdot}5/9 \pm t\,\sqrt{(s_y^2/9)}$ | **M1** | Find confidence interval (M0 if $s^2$ not $s_y^2$) |
| | $t_{8,\,0.975} = 2{\cdot}306$ or $2{\cdot}31$ | **A1** | State or use correct tabular value of t |
| | $35{\cdot}5 \pm 1{\cdot}99$ or $[33{\cdot}5, 37{\cdot}5]$ (allow $35{\cdot}5 \pm 2{\cdot}0$) | **A1** | Evaluate C.I. (either form)<br>**SC** B1 (max 1/4) for $1{\cdot}4 \pm 0{\cdot}67$ or $[0{\cdot}73, 2{\cdot}07]$ |
| | | **4** | |

**Q30.** October/November 2018

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 8(i) | $s_A^2 = (531\,000 - 5120^2/50) / 49$ and [= 6712/49]<br>$s_B^2 = (375\,135 - 3760^2/40) / 39$ [= 21695/39] | **M1** | Estimate both population variances |
| | $s_A^2 = 136.98$ and $s_B^2 = 556.28$ (to 3 s.f. throughout) | **A1** | (allow biased here: 134.24 and 542.4) |
| | EITHER: $s^2 = s_A^2/50 + s_B^2/40 = 16.65$ or $4.08^2$ | **M1A1** | EITHER: Estimate combined variance |
| | $z_{0.975} = 1.96$ | ***B1** | State or use correct tabular z (or t) value |
| | $z = (\bar{y} - \bar{x}) / s = (102.4 - 94) / s = 2.06$<br>z > tabular value [accept $H_1$] | **M1A1** | Calculate value of z (or –z)<br>(or can compare $\bar{y} - \bar{x} = 8.4$ with 8.0) |
| | so $\mu_A \neq \mu_B$ or salaries do differ (AEF) | **B1** | Correct conclusion (**FT** on z, dep *B1) |
| | OR: Assume equal [population] variances<br>$s^2 = (49\ s_A^2 + 39\ s_B^2) / 88$<br>or $(531\,000 - 5120^2/50 + 375\,135 - 3760^2/40) / 88$ | **(B1** | OR: State assumption about variances (may be in part (ii))<br>Find pooled estimate of common variance<br>(M1 A1 for $s_A^2$ and $s_B^2$ may be implied here) |
| | $= 28407/88$ or 322.8 or $17.97^2$ | **B1** | |
| | $z_{0.975} = 1.96$ | ***B1** | State or use correct tabular z (or t) value |
| | $z = 8.4 / s\sqrt{(1/50 + 1/40)} = 2.20$ | **M1A1** | Calculate value of z (or –z)<br>(or can compare $\bar{y} - \bar{x} = 8.4$ with 7.47) |
| | z > tabular value [accept $H_1$]<br>so $\mu_A \neq \mu_B$ or salaries do differ (AEF) | **B1)** | Correct conclusion (**FT** on z, dep *B1) |
| | | **8** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 8(ii) | No assumption [of normality] needed<br>since large samples or due to central limit theorem | **B1** | Valid comment on required assumptions (AEF)<br>(assumption of equal variances may earn B1 above) |
| | | **1** | |

**Q31.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 9(i) | $\bar{x}$ = 13.88/8 = 1.735 (allow 1.73 or 1.74 for this B1) | **B1** | Find sample mean |
| | $s^2$ = (24.1182 – $13.88^2$ /8) / 7<br>[ = 13/2500 or 0.0052 or $0.07211^2$ ] | **M1** | Estimate population variance<br>(allow biased here: 0.00455 or $0.06745^2$) |
| | $H_0$: $\mu = 1.7$, $H_1$: $\mu > 1.7$ (AEF) | **B1** | State hypotheses (B0 for $\bar{x}$ …) |
| | $t = (\bar{x} - 1.7) / (s/\sqrt{8}) = 1.37$ | **M1A1** | Find value of t |
| | $t_{7, 0.95} = 1.89[5]$ | **B1** | State or use correct tabular t-value<br>(or can compare $\bar{x}$ with 1.7 + 0.048 = 1.75 ) |
| | [Accept $H_0$:] Mean ht. not greater than 1.7 m (AEF) | **B1** | Consistent conclusion (**FT** on both t-values) |
| | | **7** | |
| 9(ii) | $13.88/8 \pm t \sqrt{(s^2/8)}$ | **M1** | Find confidence interval |
| | $t_{7, 0.975} = 2.36[5]$ | **A1** | State or use correct tabular value of t |
| | $1.73[5] \pm 0.06$ or [1.67, 1.80] | **A1** | Evaluate confidence interval (either form) |
| | | **3** | |

**Q32.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6 | $\bar{x} = 90.3 / 8 = 11.2875$ (to 4 s.f.) | **B1** | Find sample mean |
| | $s^2 = (1043.67 - 90.3^2/8) / 7$ | **M1** | Estimate population variance |
| | $= 19\,527 / 5600$ or 3.487 [or $1.867^2$ ] (to 4 s.f.) | **A1** | (allow biased here: 3.051 or $1.747^2$) |
| | $90.3 / 8 \pm t \sqrt{(s^2/8)}$ | **M1** | Find confidence interval |
| | $t_{7,\ 0.975} = 2.365$ (to 4 s.f.) | **A1** | State or use correct tabular value of t |
| | $11.3 \pm 1.6$ or [9.7, 12.8[5]] | **A1** | Evaluate C.I. (either form) |
| | | **6** | |

**Q33.**



| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 11B(i) | $H_0$: $\mu_x = \mu_y$ , $H_1$: $\mu_x > \mu_y$ (or in terms of $\mu_A$, $\mu_B$) | **B1** | State hypotheses (B0 for $\bar{x}$ …) |
| | $s_x^2 = (14.1775 - 10.56^2/8) / 7 = 0.03404$ and<br>$s_y^2 = (15.894 - 12.39^2/10) / 9 = 0.06031$ (to 3 s.f.) | **M1A1** | Estimate both population variances<br>(allow biased here: 0.02979 and 0.05428) |
| | $s^2 = (7\, s_x^2 + 9\, s_y^2) / 16$<br>or $(14.1775 - 10.56^2/8 + 15.894 - 12.39^2/10) / 16$<br>$= 0.78\,109 / 16$ or $0.04882$ or $0.22095^2$ (to 3 s.f.) | **M1A1** | Find pooled estimate of common variance<br>(M1 A1 for $s_x^2$ and $s_y^2$ may be implied here) |
| | $t_{16.\,0.9} = 1.337$ (to 3 s.f.) | ***B1** | State or use correct tabular t value |
| | $t = (1.32 - 1.239) / s\sqrt{(1/10 + 1/8)} = 0.773$ | **M1A1** | Calculate value of t (or –t)<br>(or can compare $\bar{y} - \bar{x} = 0.081$ with 0.140) |
| | t < tabular value<br>so claim not justified or A's not heavier than B's (AEF) | **B1** | Correct conclusion (**FT** on t, dep *B1) |
| | **SC:** $z = (1.32 - 1.239) / \sqrt{(s_x^2/8 + s_y^2/10)}$<br>$= 0.081 / \sqrt{(0.078)} = 0.799$ | **(B1** | **SC:** Implicitly taking $s_x^2$, $s_y^2$ as unequal popln. variances<br>(may also earn first B1 M1 A1) |
| | $z < 1.282$ so claim is not justified (AEF) | **B1)** | Comparison with $z_{0.9}$ and conclusion (**FT** on z; max 5/9) |
| | | **9** | |
| 11B(ii) | $\bar{x} = 1.28$ and $s^2 = 0.294 / 7$ [$= 0.042$ (or $0.205^2$)] | **M1** | Find sample mean & estimate popn. var (allow M1 if 0.294 / 8) |
| | $t = (1.28 - p) / \sqrt{(0.042/8)}$ (AEF) | **M1A1** | Find value of t |
| | $t > 1.415$ (< is A0), $p < 1.28 - 0.1025$, $p_{max} = 1.18$ | **M1A1** | Find $p_{max}$ by comparison with tabular value, here $t_{7.\,0.9}$ |
| | | **5** | |

**Q34.**

| | | | |
|---|---|---|---|
| 7 | $\bar{x} = 7{\cdot}2$ | **B1** | Find sample mean |
| | $s^2 = (542 - 72^2/10) / 9$<br>[ = 118/45 or 2·622 or $1{\cdot}619^2$ ] | **M1** | Estimate population variance<br>(allow biased here: 2·36 or $1{\cdot}536^2$) |
| | $H_0$: $\mu = 6{\cdot}2$, $H_1$: $\mu > 6{\cdot}2$ (AEF) | **B1** | State hypotheses (**B0** for $\bar{x}$ …) |
| | $t_{9,\,0.95} = 1{\cdot}83$[3] | **B1** | State or use correct tabular t-value |
| | $t = (\bar{x} - 6{\cdot}2) / (s/\sqrt{10}) = 1{\cdot}95$<br>[Accept $H_1$:] | **M1 A1** | Find value of t (or can compare $\bar{x}$ with $6{\cdot}2 + 0{\cdot}939 = 7{\cdot}14$ )<br>Consistent conclusion |
| | Claim (of mean mass increased) is justified (AEF) | **B1 FT** | (FT on both t-values) |
| | **Total:** | **7** | |

**Q35.** May/June 2017

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 9 | $H_0$: $\mu_X = \mu_Y$, $H_1$: $\mu_X \neq \mu_Y$ (AEF) | **B1** | State hypotheses (**B0** for $\bar{x}$ …) |
| | $x = 13{\cdot}4/8$ or $1{\cdot}67[5]$, $\bar{y} = 2{\cdot}02$ (all to 3 s.f.) | **B1** | Find sample means (values to 3 s.f. throughout) |
| | $s_X^2 = (24{\cdot}7 - 13{\cdot}4^2/8) / 7$<br>$= 451/1400$ or $0{\cdot}3221$ or $0{\cdot}5678^2$ and<br>$s_Y^2 = (44{\cdot}6 - 20{\cdot}2^2/10) / 9$<br>$= 949/2250$ or $0{\cdot}4218$ or $0{\cdot}6494^2$ | **M1** | Estimate or imply popln. variances<br>(allow biased here: $0{\cdot}2819$ or $0{\cdot}5309^2$)<br>(allow biased here: $0{\cdot}3796$ or $0{\cdot}6161^2$) |
| | $s^2 = (7\, s_X^2 + 9\, s_Y^2) / 16$ (AEF)<br>or $(24{\cdot}7 - 13{\cdot}4^2/8 + 44{\cdot}6 - 20{\cdot}2^2/10) / 16$ | **M1 A1** | Estimate (pooled) common variance<br>(note $s_X^2$ and $s_Y^2$ not needed explicitly) |
| | $= 6051/16\,000$ or $0{\cdot}3782$ or $0{\cdot}6150^2$ | **A1** | |
| | $t_{16,\,0.95} = 1{\cdot}746$ | ***B1** | State or use correct tabular t value |
| | [–] $t = (\bar{y} - \bar{x}) / s\sqrt{(1/8 + 1/10)} = 1{\cdot}18$ | **M1 A1** | Find value of t (or can compare $\bar{y} - \bar{x} = 0{\cdot}345$ with $0{\cdot}509$) |
| | $t < 1{\cdot}75$ so mean masses are the same (AEF) | **DB1 FT** | Correct conclusion (FT on t, dep ***B1**) |
| | SR: $Z = (\bar{y} - \bar{x}) / \sqrt{(s_X^2/8 + s_Y^2/10)}$<br>$= 0{\cdot}345 / \sqrt{(0{\cdot}078)} = 1{\cdot}20$ | **(B1)** | SR: Implicitly taking $s_X^2$, $s_Y^2$ as unequal popln. variances<br>(may also earn first **B1 B1 M1**) |
| | $Z < 1{\cdot}645$<br>so mean masses are the same (AEF) | **(B1FT)** | Comparison with $Z_{0.95}$ and conclusion (FT on $Z$)<br>(can earn at most 5/10) |
| | **Total:** | **10** | |

**Q36.**

| | | | |
|---|---|---|---|
| 6 | $(10 – 4^2/N + 102 – 8^2/2N) / (N + 2N – 2)$ (AEF) | **M1 A1** | State or find expression for pooled estimate of $\sigma^2$ (confusing biased/unbiased estimates may still earn **M1**) |
| | $112 – 48/N = 10\ (3N – 2)$, $15N^2 – 66N + 24 = 0$ | **M1 A1** | Equate to 10 and rearrange as quadratic |
| | $N = (66 \pm 54) / 30 = 4$ | **A1** | Solve quadratic for N, rejecting root 0·4 |
| | **Total:** | **5** | |

**Q37.**

| | | | |
|---|---|---|---|
| 8 | $\bar{x} = 28{\cdot}32$ and $\bar{y} = 22{\cdot}2$ | **B1** | Find both sample means |
| | $s_A^2 = (41\ 100 – 1416^2/50) / 49$ and<br>$s_B^2 = (20\ 140 – 888^2/40) / 39$ | **M1** | Estimate both population variances |
| | $s_A^2 = 20{\cdot}39$ and $s_B^2 = 164/15$ or $10{\cdot}93$ (to 3 s.f.) | **A1** | (allow biased here: 19·98 and 10·66) |
| | EITHER:<br>$s^2 = s_A^2/50 + s_B^2/40 = 0{\cdot}681$ or $0{\cdot}825^2$<br>(19·98/50 + 10·66/40 = 0·666 is M1 A0, max 8/9) | **(M1 A1** | EITHER:<br>Estimate combined variance (if biased values used wrongly here, giving zs = 1·60, only this A1 is lost) |
| | $\bar{x} – \bar{y} \pm z\,s$ | **M1** | Find confidence interval for difference |
| | $z_{0.975} = 1{\cdot}96$ | **A1** | Use appropriate tabular value |
| | $z\,s = 1{\cdot}62$ | **A1** | Evaluate semi-interval length |
| | $6{\cdot}12 \pm 1{\cdot}62$ or $[4{\cdot}50, 7{\cdot}74]$ | **A1)** | State confidence interval (in either form) |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| | OR:<br>Assume equal [population] variances<br>$s^2 = (49\ s_A^2 + 39\ s_B^2) / 88$<br>or $(41\ 100 – 1416^2/50 + 20\ 140 – 888^2/40) / 88$ | **(B1** | OR:<br>State assumption<br>Find pooled estimate of common variance<br>(**M1 A1** for $s_A^2$ and $s_B^2$ may be implied here) |
| | $= 16{\cdot}2$ or $4{\cdot}02^2$ | **B1** | |
| | $\bar{x} – \bar{y} \pm z\ s\ \sqrt{(1/50 + 1/40)}$ | **M1** | Find confidence interval for difference |
| | $z_{0.975} = 1{\cdot}96$ | **A1** | Use appropriate tabular z-value<br>(or appropriate t-value from calculator or interpolation) |
| | ($t_{88,\ 0.975} = 1{\cdot}9873$ or $1{\cdot}99$)<br>$1{\cdot}67$ ($1{\cdot}70$) | **A1** | Evaluate semi-interval length |
| | $6{\cdot}12 \pm 1{\cdot}67$ or $[4{\cdot}45, 7{\cdot}79]$<br>($6{\cdot}12 \pm 1{\cdot}70$ or $[4{\cdot}42, 7{\cdot}82]$) | **A1)** | Evaluate confidence interval (in either form) |
| | **Total:** | **9** | |

**Q38.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 10(ii) | $z = (1{\cdot}205 – 0{\cdot}91) / s = 0{\cdot}295 / s$<br>$= 2{\cdot}29[4]$ [or $2{\cdot}26[9]$] (to 3 s.f.) | **M1 A1** | Find value of z (either sign) |
| | $(z) =$ ·989[1] [or 0·988[4]] | **A1** | Find $\Phi(z)$ |
| | $100 \times (1 – 0{\cdot}989) \times 2 = 2{\cdot}2$ [or $2{\cdot}3$] (to 1 d.p.) | **M1 A1** | Find limiting value for $\alpha$, based on two-tail test<br>(**M0** for basing on one-tail test) |
| | $\alpha <$ or $\leqslant$) $2{\cdot}2$ [or $2{\cdot}3$] | **A1** | Find set of possible values of $\alpha$<br>(Treat $\alpha$ instead of $\alpha$% as misread) |
| | | **6** | |

**Q38.** October/November 2017

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 10(i) | $\bar{x} = 0{\cdot}91$, $\bar{y} = 1{\cdot}205$<br>$s_X^2 = 19{\cdot}56 / 49$ [= 489/1225 or 0·3992] and<br>$s_Y^2 = 30{\cdot}25 / 59$ [= 121/236 or 0·5127] | **B1**<br><br>**M1** | Find both sample means<br>Estimate both population variances<br>(allow biased here: 0·3912 and 0·5042) |
| | EITHER: $s^2 = s_X^2/50 + s_Y^2/60$ | **(M1** | Estimate or imply combined variance |
| | = 0·01653 or $0{\cdot}1286^2$ (to 3 s.f. throughout) | **A1** | |
| | [±] $(\bar{y} - \bar{x}) \pm z\,s$ | **M1** | Find confidence interval for difference Y – X or X – Y |
| | $z_{0.95} = 1{\cdot}645$<br>[±] 0·295 ± 0·211 (allow 0·212) | **A1** | Use appropriate tabular value<br>Evaluate confidence interval (either form) |
| | or [±] [0·084, 0·506] (allow [±] [0·083, 0·507]) | **A1)** | |
| | OR: Assume equal [population] variances<br>$s^2 = (49\,s_X^2 + 59\,s_Y^2) / 108$<br>or (19·56 + 30·25) / 108 | **(B1** | State assumption<br>Find or imply pooled estimate of common variance<br>(note $s_X^2$ and $s_Y^2$ not needed explicitly so first M1 may be implied by result) |
| | = 4981/10800 or 0·461 or $0{\cdot}679^2$ | **B1** | |
| | [±] $(\bar{y} - \bar{x}) \pm z\,s\,\sqrt{(1/50 + 1/60)}$ | **M1** | Find confidence interval for difference Y – X or X – Y |
| | $z_{0.95} = 1{\cdot}645$ | **A1** | Use appropriate tabular z-value (or t-value from calculator) |
| | [±] 0·295 ± 0·214 or [±] [0·081, 0·509] | **A1)** | Evaluate confidence interval (either form) |
| | | **7** | |

**Q39.**

| | | | | |
|---|---|---|---|---|
| **7** | Calculate sample mean: | $\bar{x} = 24{\cdot}6 / 9$ *or* 41/15 *or* 2·73[3] | **B1** | |
| | Estimate population variance: | $s^2 = (68{\cdot}5 - 24{\cdot}6^2/9) / 8$ | **M1** | |
| | (allow biased here: 0·14 *or* $0{\cdot}3742^2$) | $= 63/400$ *or* $0{\cdot}1575$ *or* $0{\cdot}3969^2$ | **B1** | |
| | State hypotheses (AEF; B0 for $\bar{x}$ ...): | $H_0: \mu = 2{\cdot}5$, $H_1: \mu > 2{\cdot}5$ | **M1 A1** | |
| | Calculate value of $t$ (to 3 s.f.): | $t = (\bar{x} - 2{\cdot}5)/(s/\sqrt{9}) = 1{\cdot}76$ | **B1** | |
| | State or use correct tabular $t$-value (to 3 s.f.) : | $t_{8,\,0.95} = 1{\cdot}86[0]$ | **M1** | |
| | (or can compare $\bar{x}$ with 2·5 + 0·246 = 2·746 ) | | | |
| | State or imply valid method for conclusion e.g.: | Accept $H_0$ if $t <$ tabular value | | |
| | Conclusion (AEF, requires both values correct): | 1·76 [ ±0·02] < 1·86 so popln. mean not greater than 2·5 | **A1** | [8] |

**Q40.**

| | | | | |
|---|---|---|---|---|
| **11B** | Find $\Sigma x$ via sample mean $\bar{x}$: | $\Sigma x = 12\,\bar{x} = 12 \times ½\,(25{\cdot}17 + 26{\cdot}83)$ | **M1 A1** | |
| | | $= 12 \times 26 = 312$ | **M1** | |
| | *EITHER:* Find estimated s.d. $s_A$, or $s_A^2$: | $t\,s_A / \sqrt{12} = ½\,(26{\cdot}83 - 25{\cdot}17)$ | | |
| | with $t = t_{11,\,0.975} = 2{\cdot}201$ | $s_A = 0{\cdot}83\,\sqrt{12} / 2{\cdot}201$ | **A1** | |
| | (to 3 s.f.) | $= 1{\cdot}306$ *or* $s_A^2 = 1{\cdot}706$ | | |
| | Find $\Sigma x^2$ from $s_A$ or $s_A^2$: | $s_A^2 = \{\,\Sigma x^2 - (\Sigma x)^2/12\,\} / 11$ | | |
| | (M0 for $s_A^2 = \{\ldots\}/12$) | $\Sigma x^2 = 11\,s_A^2 + (\Sigma x)^2/12$ | **M1 A1** | [6] |
| | | $= 8130[{\cdot}8]$ | | |
| | *OR:* Find estimated s.d. $\sigma_A$, or $\sigma_A^2$: | $t\,\sigma_A / \sqrt{11} = ½\,(26{\cdot}83 - 25{\cdot}17)$ | **(M1** | |
| | with $t = t_{11,\,0.975} = 2{\cdot}201$ | $\sigma_A = 0{\cdot}83\,\sqrt{11} / 2{\cdot}201$ | **A1** | |
| | (to 3 s.f.) | $= 1{\cdot}25$ *or* $\sigma_A^2 = 1{\cdot}564$ | | |
| | Find $\Sigma x^2$ from $\sigma_A$ or $\sigma_A^2$: | $\sigma_A^2 = \{\,\Sigma x^2 - (\Sigma x)^2/12\,\} / 12$ | | |
| | (M0 for $\sigma_A^2 = \{\ldots\}/11$) | $\Sigma x^2 = 12\,\sigma_A^2 + (\Sigma x)^2/12$ | | |
| | | $= 8130[{\cdot}8]$ | **M1 A1)** | |
| | State hypotheses (B0 for $\bar{x}_A$ …), e.g.: | $H_0: \mu_A = \mu_B$ , $H_1: \mu_A > \mu_B$ | **B1** | |
| | Estimate $B$'s popln. variance (to 3 d.p..): | $s_B^2 = (4507{\cdot}62 - 177^2/7) / 6$ | **M1** | |
| | (allow biased here: 4·578) | [ = 5·341 ] | | |
| | Find pooled estimate of common variance: | $s^2 = (11\,s_A^2 + 6\,s_B^2) / 17$ | | |
| | (note $s_B^2$ may be implied, earning M1 above) | $= (18{\cdot}77 + 32{\cdot}05) / 17$ | | |
| | | $= 2{\cdot}989$ *or* $1{\cdot}729^2$ (to 3 s.f.) | **M1 A1** | |
| | Calculate value of $t$ (or $-t$): | $t = (26 - 177/7) / s\,\sqrt{(1/12 + 1/7)}$ | | |
| | | $= (26 - 25{\cdot}29) / s\,\sqrt{(1/12 + 1/7)}$ | | |
| | (to 3 s.f.) | $= 0{\cdot}7143 / 0{\cdot}8222 = 0{\cdot}869$ | **M1 A1 *B** | |
| | State or use correct tabular $t$ value: (to 3 s.f.) | $t_{17,\,0.9} = 1{\cdot}333$ | | |
| | (or can compare $\bar{x}_A - \bar{x}_B = 0{\cdot}7143$ with 1·536) | | | |
| | Consistent conclusion (AEF, √ on $t$, dep *B1): | $t <$ tabular value | **B1√** | |
| | | so Petra's belief not supported | | |
| | | *or* wing span of $A$ not greater | | [8] |

**Q41.**

| | | | | |
|---|---|---|---|---|
| **8** | State assumption, e.g.: | Normal distribution(s) | **B1** | |
| | State hypotheses, e.g.: | $H_0$: $\mu_{Before} - \mu_{After} = 0$, | | |
| | (AEF; B0 for $\bar{x}$ … or if ambiguous) | $H_1$: $\mu_{Before} - \mu_{After} > 0$ | **B1** | |
| | Consider differences e.g. Before – After: | 0·7 3·0 –0·6 0·2 0·3 0·9 0 1·5 | **M1** | |
| | Calculate sample mean | $\bar{d} = 6 / 8 = 0{\cdot}75$ *and* | | |
| | and estimate population variance: | $s^2 = (13{\cdot}04 – 6^2/8) / 7$ | | |
| | (allow biased here:1·0675 *or* $1{\cdot}033^2$) | $= 61/50$ *or* 1·22 *or* $1{\cdot}105^2$ | **M1** | |
| | Calculate value of $t$: | $t = \bar{d}/(s/\sqrt{8}) = 1{\cdot}92$ | **M1 A1** | |
| | State or use correct tabular $t$-value: | $t_{7,\,0.975} = 2{\cdot}36$[5] | **B1** | |
| | (or compare $\bar{d}$ with $t_{7,\,0.975}\, s/\sqrt{8} = 0{\cdot}924$) | | | |
| | Consistent conclusion(AEF, √ on both $t$-values): | $H_0$:] | | |
| | | No improvement in athletes' times | **B1**√ | |
| | Wrong (hypothesis) test can earn only:B1 for assumption, B1 for hypotheses | | | [8] |

**Q42.**

| Question Number | Mark Scheme Details | | Part Mark | Total |
|---|---|---|---|---|
| **11 (b)** | State hypotheses (B0 for $\bar{x}_A$ …), e.g.: | $H_0$: $\mu_B = \mu_A$ ,$H_1$: $\mu_B > \mu_A$ | **B1** | |
| | Estimate both popn. variances: | $s_A^2 = (1910{\cdot}8 – 256^2/40) / 39$ | | |
| | *and* | $s_B^2 = (3148{\cdot}8 – 382{\cdot}9^2/50) / 49$ | **M1 A1** | |
| | (allow biased here: 6·81 *or* $2{\cdot}610^2$ | $s_A^2 = 6{\cdot}985$ *or* $454/65$ *or* $2{\cdot}643^2$ | | |
| | *and* 4·331 *or* $2{\cdot}081^2$) *and* | $_B^2 = 4{\cdot}419$ *or* $2{\cdot}102^2$ | **A1** | |
| | Estimate combined variance (may be implied): | $s^2 = s_A^2/40 + s_B^2/50$ | **M1** | |
| | (to 3 s.f) | $=0{\cdot}263$ *or* $0{\cdot}513^2$ | **A1** | |
| | Calculate value of $z$ (or $–z$): | $z = (\bar{x}_B - \bar{x}_A) / s$ | | |
| | | $= (7{\cdot}658 – 6{\cdot}4) / s$ | | |
| | (to 3 s.f.) | $= 1{\cdot}258/0{\cdot}5128 = 2{\cdot}45$ | **M1 A1** | |
| | State or use correct tabular $z$ value:(to 3 s.f.) | $z_{0.99} = 2{\cdot}326$(**SC**: allow 2·358) | ***B** | |
| | (or can compare $\bar{x}_B - \bar{x}_A = 1{\cdot}258$ with 1·19) | | | |
| | Correct conclusion (AEF, √ on $z$, dep *B1) | :$z$ > tabular value | | |
| | | so $\mu_B$ is greater than $\mu_A$ | **B1**√ | |
| | **Note**: Basing test on equal variances can earn first 4 marks only | | | [10] |
| | Calculate value of $z$ using width of interval: | $z = 1{\cdot}82 / 2s = 0{\cdot}91 / 0{\cdot}5128$ | | |
| | (allow M1 for $z = 1{\cdot}82 / s$) | $= 1{\cdot}77$[46] | **M1 A1** | |
| | Find $\Phi(z)$ and value of $\alpha$ : | $\Phi(z) = 0{\cdot}962 = 1 – ½ (1 – \alpha/100)$ | | |
| | (overlooking % is misread; | $\alpha = 100\ \{1 – 2\ (1 – 0{\cdot}962)\}$ | | |
| | M0 for $\alpha = 96{\cdot}2$ or finding $100 – \alpha$) | $= 92{\cdot}4$ | **M1 A1** | [4] |

**Q43.**

| | | | | |
|---|---|---|---|---|
| **6** | Estimate population variance:<br>(allow biased here: 0·098 or $0{\cdot}3130^2$)<br>State hypotheses (AEF; B0 for $\bar{x}$):<br>Calculate value of t (either sign; to 3 s.f.):<br>State or use correct tabular t-value (to 3 s.f.):<br>(or can compare $\bar{x}$ with 6·4 – 0·224 = 6·176 )<br>State or imply valid method for conclusion e.g.:<br>Conclusion (AEF, requires both values correct): | $s^2 = 0{\cdot}784 / 7$<br>$= 0{\cdot}112$ or 14/125 or $0{\cdot}3347^2$<br>$H_0: \mu = 6{\cdot}44$, $H_1: \mu < 6{\cdot}44$<br>$t = (6{\cdot}44 - \bar{x})/(s/\sqrt{8}) = 1{\cdot}64$<br>$t_{7,\,0.95} = 1{\cdot}89$[5]<br><br>Accept $H_0$ if t < tabular value<br>1·64 < 1·89 so<br>popln. mean not less than 6·44 | **M1**<br>**B1**<br>**M1 A1**<br>**B1**<br><br>**M1**<br><br>**A1** | [7] |

**Q44.**

| | | | | |
|---|---|---|---|---|
| **8 (i)** | EITHER: Estimate P's popln. variance (to 3 d.p.):<br>(allow biased here: 400)<br>Estimate Q's popln. variance (to 3 d.p.):<br>(allow biased here: 314)<br>Find pooled estimate of common variance:<br><br>OR: Find pooled estimate of common variance: | $s_P^2 = (1\,560\,000 - 9600^2/60) / 59$<br>[= 406·78]<br>$s_Q^2 = (1\,052\,500 - 7200^2/50) / 49$<br>[= 320·41]<br>$s^2 = (59\, s_P^2 + 49\, s_Q^2) / 108$<br>$= 367{\cdot}6$ or 368 or 9925/27$s^2 = (1\,560\,000 - 9600^2/60$<br>$s^2 = (1\,560\,000 - 9600^2/60$<br>$+ 1\,052\,500 - 7200^2/50) / 108$<br>$= 367{\cdot}6$ or 368 or 9925/27 | **M1**<br><br>**M1**<br><br>**M1 A1**<br><br>**(M3 A1)** | [4] |
| **(ii)** | Find confidence interval for the difference:<br><br>Use appropriate tabular value (to 2 d.p.):<br>Evaluate confidence interval (AEF, to 1 d.p.):<br><br>**SR** Using combined variance $s_P^2/60 + s_Q^2/50 = 13{\cdot}19$:<br>(i) M1, M1 as above; then M0 A0 (max 2/4)<br>(ii) M1 A0 for 9600/60 – 7200/50 ± z s<br>A1 for tabular value as above<br>M1 A0 for evaluating interval 16 ± 7·1 (or 7·2)<br>(max 3/5) | 9600/60 – 7200/50 [= 160 – 144]<br>$\pm\, z\, s\sqrt{(60^{-1} + 50^{-1})}$<br>$z_{0.975} = 1{\cdot}96$ or $t_{120,\,0.975} = 1{\cdot}98$<br>16 ± 7·2 or [8·8, 23·2]<br>or 16 ± 7·3 or [8·7, 23·3] | **M1 A1**<br>**A1**<br><br>**M1 A1** | [5] |

**Q45.**

| | | | | | |
|---|---|---|---|---|---|
| **6** | Equate pooled estimate of $\sigma^2$ to 12: | $(11 - 5^2/N + 160 - 10^2/10)$ / $(N + 10 - 2) = 12$ | M1 A1 | | |
| | Formulate and solve relevant quadratic eqn. for $N$: | $12N^2 - 65N + 25 = 0,\ N = 5$ | M1 A1 | 4 | **4** |

**Q46.**

| | | | | | |
|---|---|---|---|---|---|
| **7** | Find $\Sigma x$ via sample mean $\overline{x}$ : | $\Sigma x = 8\overline{x} = 8 \times ½ (1.17 + 2.03)$ $= 8 \times 1.6 = 12.8$ | M1 A1 | | |
| | Find estimate of population variance $s^2$: | $t\sqrt{(s^2/8)} = ½ (2.03 - 1.17)$ [= 0.43] | M1 | | |
| | Use of correct tabular value (1·96 leads to 23·2): | $t_{7,\,0.975} = 2.36$[5] | A1 | | |
| | (to 3 d.p.) | $s^2 = 0.2645$ *or* $32/121$ *or* $0.5143^2$ | A1 | | |
| | Find $\Sigma x^2$ from $s^2$: | $s^2 = \{\Sigma x^2 - (\Sigma x)^2/8\}/7$ | | | |
| | (M0 for $s^2 = \{\ldots\}/8$) | $\Sigma x^2 = 7 \times 0.2645 + 12.8^2/8 = 22.3$ | M1 A1 | 7 | **7** |

**Q47.**

| Question Number | Mark Scheme Details | | Part Mark | Total |
|---|---|---|---|---|
| **10** | Calculate gradient $b_1$ in $y - \bar{y} = b_1(x - \bar{x})$ : | $S_{xy} = 313.28 - 50.8 \times 56.9/10$ | | |
| | | $= 24.228$ | | |
| | | $S_{xx} = 284.16 - 50.8^2/10 = 26.096$ | | |
| | (PA –1 so max 4/5 for 0.93, giving $y = 7.48$) | $b_1 = S_{xy} / S_{xx} = 0.928$ | M1 A1 | |
| | Find $y$ when $x = 7$ from regression line of $y$ on $x$: | $y = 56.9/10 + b_1 (7 - 50.8/10)$ | M1 A1 | |
| | | $= 5.69 + 0.928 (7 - 5.08)$ | | |
| | | $[= 0.928 \times 7 + 0.976]$ | | |
| | | $y = 7.47$ (allow 7.48 or 7.5) | A1 | 5 |
| | SR If regression line of $x$ on $y$ used: | $S_{yy} = 347.59 - 56.9^2/10 = 23.829$ | | |
| | | $b_2 = S_{xy} / S_{yy} = 1.017$ | (M1) | |
| | | $7 = 50.8/10 + b_2 (y - 56.9/10)$ | (M1 A1) | |
| | | $= 1.017\, y - 0.707$ | | |
| | (can earn at most 4/5) | $y = 7.58$ (allow 7.6) | (A1) | |
| | Find differences (e.g. $y - x$) and sample mean: | 1 0.7 1.3 0.2 0.1 0.9 0.8 0.5 0.1 0.5 | | |
| | | $\bar{d} = 6{\cdot}1 / 10 = 0.61$ | M1 A1 | |
| | Estimate population variance (to 3 s.f.): | $s^2 = (5.19 - 6.1^2/10) / 9$ | | |
| | (allow biased here: 0.1469 *or* $0.3833^2$) | $= 0.1632$ *or* $0.404^2$ | B1 | |
| | State hypotheses (AEF; B0 for $\bar{x}$), e.g.: | $H_0: \mu_y - \mu_x = 0.4$, $H_1: \mu_y - \mu_x > 0.4$ | B1 | |
| | Calculate value of $t$: | $t = (\bar{d} - 0.4)/(s/\sqrt{10}) = 1.64$ | M1 A1 | |
| | State or use correct tabular $t$-value: | $t_{9,\,0.95} = 1.83$[3] | B1 | |
| | (or can compare $\bar{d}$ with 0.634) | | | |
| | Consistent conclusion (AEF, ft on both $t$-values): | [Accept $H_0$:] | | |
| | | No improvement of more than 0.4 | B1 | |
| | Wrong test can earn only B1 for hypotheses and B1 for conclusion | | 8 | |
| | | | | **13** |

**Q48.**

| | | | | |
|---|---|---|---|---|
| **8** | Calculate sample mean: | $\bar{x} = 45.29 / 7 = 6.47$ | M1 | |
| | Estimate population variance: | $s_1^2 = (295.36 – 45.29^2/7) / 6$ | | |
| | (allow biased here: 0.3334 *or* $0.5774^2$) | $= 0.389$ *or* $0.6236^2$ | M1 | |
| | State hypotheses: (AEF; B0 for $\bar{x}$) | $H_0: \mu = 6{\cdot}2$, $H_1: \mu > 6.2$ | B1 | |
| | Calculate value of $t$ (or $–t$; to 3 s.f.): | $t = (\bar{x} – 6.2)/(s_1/\sqrt{7}) = 1.15$ | M1 A1 | |
| | State or use correct tabular $t$–value (to 3 s.f.): | $t_{6,\,0.95} = 1.94$[3] | B1 | |
| | (or can compare $\bar{x}$ with $6.2 + 0.458 = 6.66$) | | | |
| | Consistent conclusion (AEF, ft on both $t$–values): | [Accept $H_0$] | | |
| | | Popn. mean not greater than 6.2 | B1✓ | |
| | Find estimate of population mean: | ½ (5.89 + 6.75) = 6.32 | M1 A1 | 7 |
| | Find estimate of population variance $s_2^2$: | $t\sqrt{(s_2^2/8)} = ½ (6.75 – 5.89)$ [= 0.43] | M1 | |
| | **SR** Allow M1 for $t\sqrt{(s_2^2/7)}$, max 4/5 | | | |
| | Use of correct tabular value: | $t_{7,\,0.975} = 2.36$[5] | A1 | |
| | (Use of 2.36 may lose final A1) | [$s_2 = 0.5143$ *or* $4\sqrt{2}/11$] | | |
| | (to 3 d.p.) | $s_2^2 = 0.2645$ *or* $32/121$ | A1 | 5 **12** |

**Q49.**

| | | | | |
|---|---|---|---|---|
| **10B** | State hypotheses (B0 for $\bar{t}_A$ …), e.g.: | $H_0: \mu_A = \mu_B$, $H_1: \mu_A < \mu_B$ | B1 | |
| | Estimate both popn. variances (to 3 d.p..): | $s_A^2 = (215.18 – 102^2/50) / 49$ | | |
| | | *and* $s_B^2 = (282.3 – 129^2/60) / 59$ | M1 A1 | |
| | (allow biased here: 0.1420 *or* $0.3768^2$ | $s_A^2 = 0.1449$ *or* $71/490$ | | |
| | | *or* $0.3807^2$ | | |
| | *and* 0.0825 *or* $0.2872^2$) | *and* $s_B^2 = 0.0839$ *or* $99/1180$ | | |
| | | *or* $0.2897^2$ | A1 | |
| | *EITHER:* Estimate combined variance: | $s^2 = s_A^2/50 + s_B^2/60$ | M1 | |
| | (to 3 s.f.) | $= 0.004296$ *or* $0.06555^2$ | A1 | |
| | Calculate value of $z$ (or $–z$): | $z = (\bar{t}_B – \bar{t}_A) / s$ | | |
| | (to 3 s.f.) | $= (2.15 – 2.04) / s = 1.68$ | M1 A1 | |
| | *OR:* Pooled estimate of common variance: | $s^2 = (49\,s_A^2 + 59\,s_B^2) / 108$ | | |
| | (note $s_A^2$ and $s_B^2$ not needed explicitly) | *or* $(215.18 – 102^2/50 +$ | | |
| | | $282.3 – 129^2/60) / 108$ | | |
| | (to 3 s.f.) | $= 0.1116$ *or* $0.334^2$ | (M1 A1) | |
| | Calculate value of $z$ (or $–z$): | $z = (\bar{t}_B – \bar{t}_A) / s\sqrt{(1/50 + 1/60)}$ | | |
| | (to 3 s.f.) | $= 0.11 / 0.06396 = 1{\cdot}72$ | (M1 A1) | |
| | State or use correct tabular $z$ value: (to 3 s.f.) | $z_{0.95} = 1.645$(allow 1.658) | *B1 | |
| | (or can compare $\bar{t}_B – \bar{t}_A = 0.11$ with 0.105) | | | |
| | Correct conclusion (AEF, ft on $z$, dep *B1): | $z >$ tabular value | | |
| | | so club $A$ takes less time | B1✓ | 10 |
| | Calculate value of $z$ (either sign): | $z = (0.11 – 0.05) / s = 0.06 / s$ | | |
| | (to 3 s.f.) | $= 0.9154$[or 0.9381] | M1 A1 | |
| | Find $\Phi(z)$ and set of possible values of $\alpha$: | $\Phi(z) = 0.820$[or 0.826] | | |
| | (M0 for $\alpha \sim 82$ or 0.82) (to 1 d.p.) | $\alpha \geqslant$ (*or* >) 18.0[or 17.4] | M1 A1 | 4 |
| | **SR** Allow B1 (max 3/4) for $\alpha < 18.0$ [or 17.4] | | | **14** |

**Q50.**

| | | | | |
|---|---|---|---|---|
| **8** | State hypotheses (B0 for $\bar{\chi}$ …):<br>Estimate both popln. variances using two samples:<br>(allow use of biased: $\sigma_{X,60}^2 = 236$ *or* $15{\cdot}36^2$)<br><br>(allow use of biased: $\sigma_{Y,50}^2 = 265$ *or* $16{\cdot}28^2$)<br>Estimate population variance for combined sample:<br><br>(allow $\sigma_{X,60}^2/60 + \sigma_{Y,50}^2/50$: $9{\cdot}233$ *or* $3{\cdot}039^2$)<br>Calculate value of $z$ (to 2 d.p., either sign):<br><br>State or use correct tabular $z$ – value (to 2 d.p.):<br>(or can compare 6 with e.g. $2{\cdot}326\ s = 7{\cdot}13$ *or* $7{\cdot}07$)<br>Correct conclusion (A.E.F, √on $z$ – values):<br>**S.R.** Assuming equal population variances:<br>Find pooled estimate of common variance $s^2$:<br><br><br><br>Calculate value of $z$ (to 2 d.p., either sign):<br><br>Tabular value; conclusion | $H_0: \mu_X = \mu_Y$, $H_1: \mu_X \neq \mu_Y$ B1<br>$S_x^2 = (626\,220 - 6060^2/60) / 59$<br>[= 240 *or* $15{\cdot}49^2$]<br>*And* $s_Y^2 = (464\,500 - 4750^2/50) / 49$<br>[= $270{\cdot}4$ *or* $16{\cdot}44^2$] M1 A1<br>$s^2 = s_X^2/60 + s_Y^2/50$<br>$= 9{\cdot}408$ *or* $3{\cdot}067^2$ M1 A1<br><br>$z = (101 - 95) / s$ M1<br>$= 6/3{\cdot}067 = 1{\cdot}96$ (*or* $1{\cdot}97$) A1<br>$z_{0.99} = 2{\cdot}326$ *or* $2{\cdot}33$ (allow $2{\cdot}36$) B1<br><br>[Accept $H_0$] Claims are the same B1√<br>Hypotheses; Explicit assumption (B1; B1)<br>: $s^2 = (626\,220 - 6060^2/60 +$<br><br>$464\,500 - 4750^2/50) / 108$ (M1 A1)<br><br>$z = 6 / s\sqrt{(1/60+1/50)} = 1{\cdot}97$ (M1 A1)<br>$= 253{\cdot}8$ *or* $15{\cdot}93^2$ (A1)<br><br>As above (B1; B1√<br>) | | **9** |

**Q51.**

| | | | | | |
|---|---|---|---|---|---|
| **B** | Estimate population variance: | $s_P^2 = (236{\cdot}0 - 42{\cdot}8^2/8)/7$ | | | |
| | (allow biased here: 0·8775 or $0{\cdot}9367^2$) | $= 351/350$ or $1{\cdot}003$ or $1{\cdot}001^2$ | M1 | | |
| | Find confidence interval (allow $z$ in place of $t$) e.g.: | $42{\cdot}8/8 \pm t\sqrt{(s_P^2/8)}$ | M1 | | |
| | Use correct tabular $t$-value: | $t_{7,\,0.975} = 2{\cdot}365$ | A1 | | |
| | Evaluate C.I. correct to 2 d.p.: | $5{\cdot}35 \pm 0{\cdot}84$ or $[4{\cdot}51, 6{\cdot}19]$ | A1 | 4 | |
| | Formulate inequality for k (or equality for $k_{\max}$): | $(5{\cdot}35 - k)/\sqrt{(s_P^2/8)} \geqslant$ [or $>$] $t$ | M1 | | |
| | Use correct tabular $t$-value: | $t_{7,\,0.9} = 1{\cdot}415$ | A1 | | |
| | Solve for $k_{\max}$ (A0 if = or $\leqslant$ was used for $k$ above): | $5{\cdot}35 - k \geqslant 0{\cdot}50$, $k_{\max} = 4{\cdot}85$ | A1 | 3 | |
| | State hypotheses (B0 for $\bar{x}$ …), e.g.: | $H_0$: $\mu_P = \mu_Q$, $H_1$: $\mu_P > \mu_Q$ | B1 | | |
| | State assumption (A.E.F.): | Normal distns. for [$P$ and] $Q$ | | | |
| | | *and* equal variances | B1 | | |
| | Estimate (pooled) common variance: | $s^2 = (7 \times 1{\cdot}003 + 11 \times 1{\cdot}962)/18$ | | | |
| | | $= 1{\cdot}589$ or $1{\cdot}261^2$ | M1 A1 | | |
| | Calculate value of $t$ (to 3 s.f.): | $t = (5{\cdot}35 - 4{\cdot}60)/(s\sqrt{(1/8 + 1/12)})$ | | | |
| | | $= 1{\cdot}30$ | M1 A1 | | |
| | Correct conclusion (A.E.F., $\surd$ on $t$): | $t < t_{18,\,0.9} = 1{\cdot}33$ so Q's mean is not less than P's | B1$\surd$ | 7 | **14** |

**Q52.**

| | | | |
|---|---|---|---|
| **6** | Consider differences e.g. before – after: 8 3 –4 2 –11 3 17 10 6 0 M1<br>Calculate sample mean $\bar{d} = 34/10 = 3{\cdot}4$ *and*<br>and estimate population variance: $s^2 = (648 - 34^2/10)/9$<br>(allow biased here: $53{\cdot}24$ *or* $7{\cdot}297^2$) = /45 *or* $59{\cdot}16$ *or* $7{\cdot}691^2$ M1<br>State hypotheses (A.E.F.; B0 for $\bar{x}$), e.g.: $_0$: $\mu_B - \mu_A = 0$, $H_1$: $\mu_B - \mu_A > 0$ B1<br>Calculate value of $t$: $t = \bar{d}/(s/\sqrt{10}) = 1{\cdot}398$ *or* $1{\cdot}4$ M1 A1<br>State or use correct tabular $t$-value: $t_{9,\,0.9} = 1{\cdot}38[3]$ B1<br>(or can compare $\bar{d}$ with $3{\cdot}36[4]$)<br>Consistent conclusion (A.E.F, √ on both $t$–values): [Reject $H_0$:]<br>Hours of absence have decreased B1√<br>Wrong test can earn only B1 for hypotheses<br>and B1 for conclusion | 7 | **7** |

**Q53.**

| | | | |
|---|---|---|---|
| **6** | Estimate population variance for combined sample:<br>$s^2 = \frac{s_X^{\,2}}{50} + \frac{s_Y^{\,2}}{60}$<br>$= \frac{1391}{1500}$ *or* $0{\cdot}9273$ *or* $0{\cdot}9630^2$ M1<br>Calculate value of $z$ (to 2 d.p., either sign): $z = \frac{1.8}{s} = 1{\cdot}869$ M1 A1<br>Find $\Phi(z)$ and set of possible values of $\alpha$ (to 1 d.p.):<br>$\Phi(z) = 0{\cdot}9692$ [*or* 96·92%]<br>(M1 A0 for $\alpha \leqslant 3{\cdot}1$ *or* $\alpha > 93{\cdot}8$) $\alpha \leqslant$ (*or* <) $6{\cdot}2$ (allow $6{\cdot}1$) M1 A1 | 5 | **[5]** |

# 3. Chi Squared Tests

**Q1.**

**4** A scientist is investigating the numbers of a particular type of butterfly in a certain region. He claims that the numbers of these butterflies found per square metre can be modelled by a Poisson distribution with mean 2.5. He takes a random sample of 120 areas, each of one square metre, and counts the number of these butterflies in each of these areas. The following table shows the observed frequencies together with some of the expected frequencies using the scientist's Poisson distribution.

| Number per square metre | 0 | 1 | 2 | 3 | 4 | 5 | 6 | $\geqslant 7$ |
|---|---|---|---|---|---|---|---|---|
| Observed frequency | 12 | 20 | 36 | 32 | 13 | 6 | 1 | 0 |
| Expected frequency | 9.85 | 24.63 | 30.78 | 25.65 | $p$ | 8.02 | 3.34 | $q$ |

**(a)** Find the values of $p$ and $q$, correct to 2 decimal places. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Carry out a goodness of fit test, at the 10% significance level, to test the scientist's claim. [6]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2022

**Q2.**

**2** A scientist is investigating the size of shells at various beach locations. She selects four beach locations and takes a random sample of shells from each of these beaches. She classifies each shell as large or small. Her results are summarised in the following table.

| | | Beach location | | | | |
|---|---|---|---|---|---|---|
| | | *A* | *B* | *C* | *D* | Total |
| Size of shell | Large | 68 | 69 | 96 | 81 | 314 |
| | Small | 28 | 55 | 64 | 39 | 186 |
| | Total | 96 | 124 | 160 | 120 | 500 |

Test, at the 10% significance level, whether the size of shell is independent of the beach location. [7]

**Q3.**

**2** A driving school employs four instructors to prepare people for their driving test. The allocation of people to instructors is random. For each of the instructors, the following table gives the number of people who passed and the number who failed their driving test last year.

| | Instructor *A* | Instructor *B* | Instructor *C* | Instructor *D* | Total |
|---|---|---|---|---|---|
| Pass | 72 | 42 | 52 | 68 | 234 |
| Fail | 33 | 34 | 41 | 58 | 166 |
| Total | 105 | 76 | 93 | 126 | 400 |

Test at the 10% significance level whether success in the driving test is independent of the instructor. [7]

**Q4.**

**5** Chai packs china mugs into cardboard boxes. Chai's manager suspects that breakages occur at random times and that the number of breakages may follow a Poisson distribution. He takes a small sample of observations and finds that the number of breakages in a one-hour period has a mean of 2.4 and a standard deviation of 1.5.

**(a)** Explain how this information tends to support the manager's suspicion. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The manager now takes a larger sample and claims that the numbers of breakages in a one-hour period follow a Poisson distribution. The numbers of breakages in a random sample of 180 one-hour periods are summarised in the following table.

| Number of breakages | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 or more |
|---|---|---|---|---|---|---|---|---|
| Frequency | 21 | 33 | 46 | 31 | 23 | 16 | 10 | 0 |

The mean number of breakages calculated from this sample is 2.5.

**(b)** Use the data from this larger sample to carry out a goodness of fit test, at the 10% significance level, to test the claim. [8]

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

**Q5.**

**3** A supermarket sells pears in packs of 8. Some of the pears in a pack may not be ripe, and the supermarket manager claims that the number of unripe pears in a pack can be modelled by the distribution B(8, 0.15).

A random sample of 150 packs was selected and the number of unripe pears in each pack was recorded. The following table shows the observed frequencies together with some of the expected frequencies using the manager's binomial distribution.

| Number of unripe pears per pack | 0 | 1 | 2 | 3 | 4 | 5 | $\geqslant 6$ |
|---|---|---|---|---|---|---|---|
| Observed frequency | 35 | 48 | 43 | 15 | 6 | 3 | 0 |
| Expected frequency | 40.874 | $p$ | 35.641 | 12.579 | 2.775 | 0.392 | $q$ |

**(a)** Find the values of $p$ and $q$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Carry out a goodness of fit test, at the 5% significance level, to test whether the manager's claim is justified. [6]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

**Q6.**

**2** It is claimed that the heights of a particular age group of boys follow a normal distribution with mean 125 cm and standard deviation 12 cm. Observations for a randomly chosen group of 60 boys in this age group are summarised in the following table. The table also gives the expected frequencies, correct to 2 decimal places, based on the normal distribution with mean 125 cm and standard deviation 12 cm.

| Height, $x$ cm | $x < 100$ | $100 \leqslant x < 110$ | $110 \leqslant x < 120$ | $120 \leqslant x < 130$ | $130 \leqslant x < 140$ | $x \geqslant 140$ |
|---|---|---|---|---|---|---|
| Observed frequency | 0 | 3 | 15 | 23 | 11 | 8 |
| Expected frequency | 1.12 | 5.22 | 13.97 | 19.38 | 13.97 | 6.34 |

**(a)** Show how the expected frequency for $130 \leqslant x < 140$ is obtained. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Carry out a goodness of fit test, at the 5% significance level, to determine whether the claim is supported by the data. [6]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

**Q7.**

**1** Two randomly selected groups of students, with similar ranges of abilities, take the same examination in different rooms. One group of 140 students takes the examination with background music playing. The other group of 210 students takes the examination in silence. Each student is awarded a grade for their performance in the examination and the numbers from each group gaining each grade are shown in the following table.

| | Grade awarded | | |
|---|---|---|---|
| | A | B | C |
| Background music | 49 | 51 | 40 |
| Silence | 93 | 68 | 49 |

Test at the 10% significance level whether grades awarded are independent of whether background music is playing during the examination. [6]

**Q8.**

**1** Young children are learning to read using two different reading schemes, *A* and *B*. The standards achieved are measured against the national average standard achieved and classified as above average, average or below average. For two randomly chosen groups of young children, the numbers in each category are shown in the table.

| | Standard achieved | | |
|---|---|---|---|
| | Above average | Average | Below average |
| Scheme *A* | 31 | 35 | 22 |
| Scheme *B* | 19 | 50 | 43 |

Test at the 5% significance level whether standard achieved is independent of the reading scheme used. [6]

**Q9.**

**3** Apples are sold in bags of 5. Based on her previous experience, Freya claims that the probability of any apple weighing more than 100 grams is 0.35, independently of other apples in the bag.

The apples in a random sample of 150 bags are checked and the number, $x$, in each bag weighing more than 100 grams is recorded. The results are shown in the following table.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Frequency | 12 | 39 | 46 | 37 | 12 | 4 |

Carry out a goodness of fit test at the 5% significance level and hence comment on Freya's claim. [7]

**Q10.**

**3** A random sample of 200 observations of the continuous random variable $X$ was taken and the values are summarised in the following table.

| Interval | $0 \leqslant x < 0.5$ | $0.5 \leqslant x < 1$ | $1 \leqslant x < 1.5$ | $1.5 \leqslant x < 2$ | $2 \leqslant x < 2.5$ | $2.5 \leqslant x < 3$ |
|---|---|---|---|---|---|---|
| Observed frequency | 5 | 23 | 40 | 41 | 46 | 45 |

It is required to test the goodness of fit of the distribution with probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{9}x(4-x) & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

Most of the relevant expected frequencies, correct to 2 decimal places, are given in the following table.

| Interval | $0 \leqslant x < 0.5$ | $0.5 \leqslant x < 1$ | $1 \leqslant x < 1.5$ | $1.5 \leqslant x < 2$ | $2 \leqslant x < 2.5$ | $2.5 \leqslant x < 3$ |
|---|---|---|---|---|---|---|
| Expected frequency | $p$ | $q$ | 37.96 | 43.52 | 43.52 | 37.96 |

**(a)** Show that $p = 10.19$ and find the value of $q$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2020

**(b)** Carry out a goodness of fit test, at the 5% significance level, to test whether f is a satisfactory model for the data. [4]

**Q11.**

**9** A random sample of 50 observations of the continuous random variable $X$ was taken and the values are summarised in the following table.

| Interval | $0 \leqslant x < 0.8$ | $0.8 \leqslant x < 1.6$ | $1.6 \leqslant x < 2.4$ | $2.4 \leqslant x < 3.2$ | $3.2 \leqslant x < 4$ |
|---|---|---|---|---|---|
| Observed frequency | 18 | 16 | 8 | 6 | 2 |

It is required to test the goodness of fit of the distribution with probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{3}{16}(4-x)^{\frac{1}{2}} & 0 \leqslant x < 4, \\ 0 & \text{otherwise.} \end{cases}$$

The relevant expected frequencies, correct to 2 decimal places, are given in the following table.

| Interval | $0 \leqslant x < 0.8$ | $0.8 \leqslant x < 1.6$ | $1.6 \leqslant x < 2.4$ | $2.4 \leqslant x < 3.2$ | $3.2 \leqslant x < 4$ |
|---|---|---|---|---|---|
| Expected frequency | 14.22 | 12.54 | 10.59 | 8.18 | 4.47 |

**(i)** Show how the expected frequency for $1.6 \leqslant x < 2.4$ is obtained. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2019

**(ii)** Carry out a goodness of fit test at the 5% significance level. [7]

**Q12.**

**8** Two salesmen, *A* and *B*, work at a company that arranges different types of holidays: self-catering, hotel and cruise. The table shows, for a random sample of 150 holidays, the number of each type arranged by each salesman.

| | | Type of holiday | | |
|---|---|---|---|---|
| | | Self-catering | Hotel | Cruise |
| Salesman | *A* | 25 | 38 | 21 |
| | *B* | 28 | 21 | 17 |

Test at the 10% significance level whether the type of holiday arranged is independent of the salesman. [8]

**Q13.**

The number of puncture repairs carried out each week by a small repair shop is recorded over a period of 40 weeks. The results are shown in the following table.

| Number of repairs in a week | 0 | 1 | 2 | 3 | 4 | 5 | $\geqslant 6$ |
|---|---|---|---|---|---|---|---|
| Number of weeks | 6 | 15 | 9 | 6 | 3 | 1 | 0 |

**(i)** Calculate the mean and variance for the number of repairs in a week and comment on the possible suitability of a Poisson distribution to model the data. [3]

.....

.....

.....

.....

.....

.....

.....

Records over a longer period of time indicate that the mean number of repairs in a week is 1.6. The following table shows some of the expected frequencies, correct to 3 decimal places, for a period of 40 weeks using a Poisson distribution with mean 1.6.

| Number of repairs in a week | 0 | 1 | 2 | 3 | 4 | 5 | $\geqslant 6$ |
|---|---|---|---|---|---|---|---|
| Expected frequency | 8.076 | 12.921 | 10.337 | 5.513 | 2.205 | $a$ | $b$ |

**(ii)** Show that $a = 0.706$ and find the value of the constant $b$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2019

.....

.....

.....

.....

.....

.....

**(iii)** Carry out a goodness of fit test of a Poisson distribution with mean 1.6, using a 10% significance level. [8]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q14.**

**8** A manufacturer produces three types of car: hatchbacks, saloons and estates. Each type of car is available in one of three colours: silver, blue and red. The manufacturer wants to know whether the popularity of the colour of the car is related to the type of car. A random sample of 300 cars chosen by customers gives the information summarised in the following table.

| | | Colour of car | | |
|---|---|---|---|---|
| | | Silver | Blue | Red |
| Type of car | Hatchback | 53 | 36 | 41 |
| | Saloon | 29 | 40 | 31 |
| | Estate | 28 | 24 | 18 |

Test at the 10% significance level whether the colour of car chosen by customers is independent of the type of car. [8]

**Q15.**

11 A scientist carries out an experiment to investigate the quantity $X$, which takes the values 0, 1, 2, 3, 4, 5 or 6. He believes that the values taken by $X$ follow a binomial distribution. He conducts 250 trials. His results are summarised in the following table.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Observed frequency | 22 | 83 | 72 | 53 | 17 | 3 | 0 |

**(i)** Show that unbiased estimates of the mean and variance for these results are 1.876 and 1.266 respectively, correct to 3 decimal places. By evaluating the mean and variance of the distribution B(6, 0.313), explain why $X$ could have this distribution. [4]

.....

.....

.....

.....

.....

.....

.....

.....

The expected frequencies corresponding to the distribution B(6, 0.313) are shown in the following table.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Observed frequency | 22 | 83 | 72 | 53 | 17 | 3 | 0 |
| Expected frequency | 26.3 | 71.9 | 81.8 | 49.7 | 17.0 | 3.1 | 0.2 |

**(ii)** Show how the expected frequency for $x = 4$ is calculated. [2]

.....

.....

.....

.....

.....

.....

© UCLES 2018

**(iii)** Test at the 5% significance level whether the scientist's belief is correct. [8]

**Q16.**

**11** A machine is used to produce metal rods. When the machine is working efficiently, the lengths, $x$ cm, of the rods have a normal distribution with mean 150 cm and standard deviation 1.2 cm. The machine is checked regularly by taking random samples of 200 rods. The latest results are shown in the following table.

| Interval | $146 \leqslant x < 147$ | $147 \leqslant x < 148$ | $148 \leqslant x < 149$ | $149 \leqslant x < 150$ |
|---|---|---|---|---|
| Observed frequency | 1 | 2 | 23 | 52 |

| $150 \leqslant x < 151$ | $151 \leqslant x < 152$ | $152 \leqslant x < 153$ | $153 \leqslant x < 154$ |
|---|---|---|---|
| 69 | 36 | 15 | 2 |

As a first check, the sample is used to calculate an estimate for the mean.

**(i)** Show that an estimate for the mean from this sample is close to 150 cm. [2]

.....

.....

.....

.....

As a second check, the results are tested for goodness of fit of the normal distribution with mean 150 cm and standard deviation 1.2 cm. The relevant expected frequencies, found using the normal distribution function given in the List of Formulae (MF10), are shown in the following table.

| Interval | $x < 147$ | $147 \leqslant x < 148$ | $148 \leqslant x < 149$ | $149 \leqslant x < 150$ |
|---|---|---|---|---|
| Observed frequency | 1 | 2 | 23 | 52 |
| Expected frequency | 1.24 | 8.32 | 30.94 | 59.50 |

| $150 \leqslant x < 151$ | $151 \leqslant x < 152$ | $152 \leqslant x < 153$ | $153 \leqslant x$ |
|---|---|---|---|
| 69 | 36 | 15 | 2 |
| 59.50 | 30.94 | 8.32 | 1.24 |

**(ii)** Show how the expected frequency for $151 \leqslant x < 152$ is obtained. [3]

.....

.....

.....

.....

.....

© UCLES 2018

**(iii)** Test, at the 5% significance level, the goodness of fit of the normal distribution to the results. [7]

**Q17.**

**10** The number of accidents, $x$, that occur each day on a motorway are recorded over a period of 40 days. The results are shown in the following table.

| Number of accidents | 0 | 1 | 2 | 3 | 4 | 5 | 6 | ⩾7 |
|---|---|---|---|---|---|---|---|---|
| Observed frequency | 3 | 5 | 8 | 10 | 5 | 7 | 2 | 0 |

**(i)** Show that the mean number of accidents each day is 2.95 and calculate the variance for this sample. Explain why these values suggest that a Poisson distribution might fit the data. [3]

.....

.....

.....

.....

.....

.....

.....

.....

A Poisson distribution with mean 2.95, as found from the data, is used to calculate the expected frequencies, correct to 2 decimal places. The results are shown in the following table.

| Number of accidents | 0 | 1 | 2 | 3 | 4 | 5 | 6 | ⩾7 |
|---|---|---|---|---|---|---|---|---|
| Observed frequency | 3 | 5 | 8 | 10 | 5 | 7 | 2 | 0 |
| Expected frequency | 2.09 | 6.18 | 9.11 | 8.96 | 6.61 | 3.90 | 1.92 | 1.23 |

**(ii)** Show how the expected frequency of 6.61 for $x = 4$ is obtained. [2]

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2018

**(iii)** Test at the 5% significance level the goodness of fit of this Poisson distribution to the data. [7]

## Q18.

**11** A shop is supplied with large quantities of plant pots in packs of six. These pots can be damaged easily if they are not packed carefully. The manager of the shop is a statistician and he believes that the number of damaged pots in a pack of six has a binomial distribution. He chooses a random sample of 250 packs and records the numbers of damaged pots per pack. His results are shown in the following table.

| Number of damaged pots per pack ($x$) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Frequency | 48 | 69 | 78 | 32 | 22 | 1 | 0 |

**(i)** Show that the mean number of damaged pots per pack in this sample is 1.656. [1]

.....

.....

.....

.....

.....

The following table shows some of the expected frequencies, correct to 2 decimal places, using an appropriate binomial distribution.

| Number of damaged pots per pack ($x$) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Expected frequency | 36.01 | 82.36 | $a$ | 39.89 | $b$ | 1.74 | 0.11 |

**(ii)** Find the values of $a$ and $b$, correct to 2 decimal places [5]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(iii)** Use a goodness-of-fit test at the 1% significance level to determine whether the manager's belief is justified. [8]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2017

**Q19.**

**10** Roberto owns a small hotel and offers accommodation to guests. Over a period of 100 nights, the numbers of rooms, $x$, that are occupied each night at Roberto's hotel and the corresponding frequencies are shown in the following table.

| Number of rooms occupied ($x$) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | $\geqslant 7$ |
|---|---|---|---|---|---|---|---|---|
| Number of nights | 4 | 9 | 18 | 26 | 20 | 16 | 7 | 0 |

**(i)** Show that the mean number of rooms that are occupied each night is 3.25. [1]

.....

.....

.....

The following table shows most of the corresponding expected frequencies, correct to 2 decimal places, using a Poisson distribution with mean 3.25.

| Number of rooms occupied ($x$) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | $\geqslant 7$ |
|---|---|---|---|---|---|---|---|---|
| Observed frequency | 4 | 9 | 18 | 26 | 20 | 16 | 7 | 0 |
| Expected frequency | 3.88 | 12.60 | 20.48 | 22.18 | 18.02 | 11.72 | | |

**(ii)** Show how the expected value of 22.18, for $x = 3$, is obtained and find the expected values for $x = 6$ and for $x \geqslant 7$. [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2017

**(iii)** Use a goodness-of-fit test at the 5% significance level to determine whether the Poisson distribution is a suitable model for the number of rooms occupied each night at Roberto's hotel. [7]

**Q20.**

**8** Members of a Statistics club are voting to elect a new president of the club. Members must choose to vote either by post or by text or by email. The method of voting chosen by a random sample of 60 male members and 40 female members is given in the following table.

| | | Method of voting | | |
|---|---|---|---|---|
| | | Post | Text | Email |
| Gender | Male | 10 | 12 | 38 |
| | Female | 5 | 21 | 14 |

Test, at the 1% significance level, whether there is an association between method of voting and gender. [8]

**Q21.**

**9**

| | Mint | Strawberry | Chocolate | Fudge |
|---|---|---|---|---|
| Male | 25 | 12 | 30 | 13 |
| Female | 16 | 22 | 25 | 7 |

Test, at the 10% significance level, whether there is a difference between ice cream flavour preferences of male and female customers. [9]

© UCLES 2016

**[Turn over**

**Q22.**

**9** The number of visitors arriving at an art exhibition is recorded for each 10-minute period of time during the ten hours that it is open on a particular day. The results are as follows.

| Number of visitors in a 10-minute period | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | ⩾9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of 10-minute periods | 2 | 2 | 12 | 8 | 11 | 13 | 4 | 7 | 1 | 0 |

**(i)** Calculate the mean and variance for this sample and explain whether your answers support a suggestion that a Poisson distribution might be a suitable model for the number of visitors in a 10-minute period. [3]

**(ii)** Use an appropriate Poisson distribution to find the two expected frequencies missing from the following table. [2]

| Number of visitors in a 10-minute period | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | ⩾9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Expected number of 10-minute periods | 1.10 | | 8.79 | | 11.72 | 9.38 | 6.25 | 3.57 | 1.79 | 1.28 |

**(iii)** Test, at the 10% significance level, the goodness of fit of this Poisson distribution to the data. [8]

© UCLES 2016

**[Turn over**

**Q23.**

**11** Each of 200 identically biased dice is thrown repeatedly until an even number is obtained. The number of throws, $x$, needed is recorded and the results are summarised in the following table.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | $\geqslant 7$ |
|---|---|---|---|---|---|---|---|
| Frequency | 126 | 43 | 22 | 3 | 5 | 1 | 0 |

State a type of distribution that could be used to fit the data given in the table above. [1]

Fit a distribution of this type in which the probability of throwing an even number for each die is 0.6 and carry out a goodness of fit test at the 5% significance level. [8]

For each of these dice, it is known that the probability of obtaining a 6 when it is thrown is 0.25. Ten of these dice are each thrown 5 times. Find the probability that at least one 6 is obtained on exactly 4 of the 10 dice. [5]

© UCLES 2015

**Q24.**

**6** The reliability of the broadband connection received from two suppliers, $A$ and $B$, is classified as good, fair or poor by a random sample of householders. The information collected is summarised in the following table.

| | | Reliability | | |
|---|---|---|---|---|
| | | Good | Fair | Poor |
| Supplier | $A$ | 65 | 63 | 33 |
| | $B$ | 51 | 44 | 44 |

Test, at the 5% significance level, whether reliability is independent of supplier. [8]

© UCLES 2015

**Q25.**

**9** A random sample of 200 observations of the continuous random variable $X$ was taken and the values are summarised in the following table.

| Interval | $1 \leqslant x < 2$ | $2 \leqslant x < 3$ | $3 \leqslant x < 4$ | $4 \leqslant x < 5$ | $5 \leqslant x < 6$ | $6 \leqslant x < 7$ | $7 \leqslant x < 8$ |
|---|---|---|---|---|---|---|---|
| Observed frequency | 63 | 45 | 32 | 25 | 22 | 7 | 6 |

It is required to test the goodness of fit of the distribution with probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \dfrac{1}{x \ln 8} & 1 \leqslant x < 8, \\ 0 & \text{otherwise.} \end{cases}$$

The relevant expected frequencies, correct to 2 decimal places, are given in the following table.

| Interval | $1 \leqslant x < 2$ | $2 \leqslant x < 3$ | $3 \leqslant x < 4$ | $4 \leqslant x < 5$ | $5 \leqslant x < 6$ | $6 \leqslant x < 7$ | $7 \leqslant x < 8$ |
|---|---|---|---|---|---|---|---|
| Expected frequency | 66.67 | $p$ | 27.67 | $q$ | 17.54 | 14.83 | 12.84 |

Show that $p = 39.00$, correct to 2 decimal places, and find the value of $q$. [4]

Carry out a goodness of fit test at the 5% significance level. [6]

© UCLES 2014

# MS
# Chi-Squared Tests

**Q1.** May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4(a) | $p = 16.03$ | **B1** | |
| | $q = 1.70$ | **B1** | Condone 1.7 |
| | | **2** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4(b) | <table><tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>$\geqslant 7$</td></tr><tr><td>12</td><td>20</td><td>36</td><td>32</td><td>13</td><td>6</td><td>1</td><td>0</td></tr><tr><td>9.85</td><td>24.63</td><td>30.78</td><td>25.65</td><td>***16.03***</td><td>8.02</td><td>3.34</td><td>***1.70***</td></tr></table> Combine last two columns: O value = 1, E value = 5.04 | **M1** | Add last 2 columns (or 3 columns: 13.06). |
| | Calculate values of $\frac{(O-E)^2}{E}$ :<br>0.4693 0.8704 0.8853 1.5720<br>0.5727 0.5088 3.2384 | **M1** | Calculate values. |
| | Test statistic = 8.12 | **A1** | |
| | $H_0$: data fits a Poisson distribution with mean 2.5<br>$H_1$: data does not fit a Poisson distribution with mean 2.5 | **B1** | Need data and distribution.<br>e.g Data fits the given distribution.<br>Number of butterflies per square metre fits Po(2.5). |
| | Critical value of chi-squared = 10.64<br>Compare: '8.12' < 10.64, accept $H_0$ | **M1** | Compare their test statistic with 10.64.<br>Note: allow 9.236 if 3 columns combined<br>or 12.02 if none combined. |
| | There is sufficient evidence to support the scientist's claim / there is sufficient evidence to suggest that the data fits a Poisson distribution with mean 2.5 | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **6** | |

**Q2.**

May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2 | $H_0$ : the size of shell is independent of the location<br>$H_1$ : the size of shell is not independent of the location | **B1** | |
| | <table><tr><td>O</td><td>68</td><td>69</td><td>96</td><td>81</td></tr><tr><td>E</td><td>60.29</td><td>77.87</td><td>100.48</td><td>75.36</td></tr></table><br><table><tr><td>O</td><td>28</td><td>55</td><td>64</td><td>39</td></tr><tr><td>E</td><td>35.71</td><td>46.13</td><td>59.52</td><td>44.64</td></tr></table> | **M1 A1** | Attempt at calculation of E values. |
| | Calculate values of $\frac{(O-E)^2}{E}$ :<br>0.9859 1.0104 0.1997 0.4221<br>1.6646 1.7055 0.3372 0.7126 | **M1** | Attempt at Chi-squared contributions.<br>Accuracy of $\pm 0.1$. |
| | Test statistic = 7.04 | **A1** | Correct total to 3sf. |
| | Tabular value for 3df = 6.251<br>7.04 > 6.251<br>Reject $H_o$ | **M1** | Compare with 6.251 and correct FT conclusion.<br>May be implied by FT correct conclusion in words. |
| | Insufficient evidence to suggest size of shell is independent of location/Evidence suggests that size of shell is dependent on (beach) location | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **7** | |

**Q3.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2 | <table><tr><td>72 (61.425)</td><td>42 (44.46)</td><td>52 (54.405)</td><td>68 (73.71)</td></tr><tr><td>33 (43.575)</td><td>34 (31.54)</td><td>41 (38.595)</td><td>58 (52.29)</td></tr></table> | **M1** | At least 3 correct to 4sf. |
| | Calculates E values (in brackets above) | **A1** | All correct to 4sf (if 6.037 not found). |
| | Test value: $\sum\frac{(O-E)^2}{E}$ = 1.8206 + 0.1361 + 0.1063 +0.4423 + 2.5664 + 0.1919 + 0.1499 + 0.6235 | **M1** | Correct formula used. |
| | = 6.037 | **A1** | Correct to 3 sf. |
| | $H_0$: (driving) test success is independent of instructor | **B1** | |
| | Tabular value: 3 df 10% = 6.251<br>6.04 < 6.251<br>Accept $H_0$ | **M1** | Compare with correct tabular value and conclusion. |
| | Sufficient evidence to suggest that (driving) test success is independent of instructor | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used.<br>Allow $\pm 1$ in the third significant figure of test value. |
| | | **7** | |

**Q4.** May/June 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5(a) | (Mean of distribution = 2.4)<br>Variance = 2.25 | **M1** | Allow $1.5^2$. |
| | Mean approx. equal to variance (so Poisson might be suitable) | **A1** | Must have 'approximately' oe, not equality.<br>Must have 2.25. |
| | | **2** | |
| 5(b) | Po(2.5) leads to frequencies: | **M1** | At least 3 correct to 3sf. |
| | 14.77[5], 36.938, 46.173, 38.477, 24.048, 12.024, 5.010, 2.554 | **A1** | All correct to 4sf. |
| | Combine last two values: 7.564 | **M1** | FT their table values if final figure is less than 5. |
| | $\sum \frac{(O-E)^2}{E} = 2.6227 + 0.4198 + 0.00065 + 1.4529 + 0.04570 + 1.3148 + 0.7845$ | **M1** | Apply correct formula to their frequencies. |
| | 6.64 | **A1** | Correct to 3sf. |
| 5(b) | $H_0$: distribution fits the data | **B1** | Must mention data and distribution . |
| | Tabular value, 5 degrees of freedom, is 9.236<br>'6.64' < 9.236, so accept $H_0$ | **M1** | Compare with correct tabular value and conclusion. |
| | Insufficient evidence to show that data does not follow a Poisson distribution | **A1** | Correct conclusion, in context, following correct work.<br>Level of uncertainty in language is used.<br>Allow ±1 difference in third significant figure. |
| | | **8** | |

**Q5.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(a) | <table><tr><td>$x$</td><td>**0**</td><td>1</td><td></td><td>3</td><td>4</td><td>5</td><td>6 or more</td></tr><tr><td>O</td><td>3</td><td>48</td><td>43</td><td>15</td><td>6</td><td>3</td><td>0</td></tr><tr><td>E</td><td>40.874</td><td>57.704</td><td>35.641</td><td>12.579</td><td>2.775</td><td>0.392</td><td>0.0352</td></tr></table> $p = 57.704$ | **B1** | 3 dp or better. |
| | $q = 0.035$ or $0.036$ | **B1** | Accept numbers which round to 0.035 or 0.036. |
| | | **2** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(b) | Combine frequencies less than 5: last 4 columns give 24 / 15.7812 | **M1** | Allow last 2 or 3 columns combined for this M1. |
| | $\sum \frac{(O-E)^2}{E} = 0.8444 + 1.6319 + 1.5199 + 4.2803$ | **M1** | |
| | 8.28 | **A1** | Accept 8.27 – 8.28.<br>**SC:** 25.60, if values not combined, scores M1A0.<br>**SC:** 14.96 if last 3 combined, scores M1A0.<br>**SC:** if last 2 combined, scores M1A0. |
| | $H_0$: B(8, 0.15) fits the data | **B1** | Must mention distribution and data. |
| | 3 degrees of freedom, tabular value = 7.815.<br>8.28 > 7.815<br>Reject $H_0$. | **M1** | Compare their value with 7.815 and correct FT conclusion<br>3 columns combined compared with 9.488 and correct conclusion |
| | Insufficient evidence to support manager's claim. | **A1** | Correct conclusion, in context, following correct work.<br>Level of uncertainty in language is used. |
| | | **6** | |

**Q6.** October/November 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2(a) | $P(130 \leqslant x < 140) = P\left(\frac{130-125}{12} \leqslant Z < \frac{140-125}{12}\right)$<br>$P(0.4167 < Z < 1.25)$ | **M1** | Must be seen.<br>Must see 0.4167 and 1.25, may be implied by 0.8944 and 0.6616<br>Accept 0.417. |
| | $0.8944 - 0.6616 = 0.2328$<br>Multiply by 60 leads to 13.97 | **A1** | AG<br>Accept 0.6615. Need to see 0.2328 or 0.2329 or 13.968 or $(0.8944 - 0.6616) \times 60$. |
| | | **2** | |
| 2(b) | <table><tr><td>0</td><td>3</td><td>15</td><td>23</td><td>11</td><td>8</td></tr><tr><td>1.12</td><td>5.22</td><td>13.97</td><td>19.38</td><td>13.97</td><td>6.34</td></tr></table> | **M1** | Combine first two values. |
| | Test stat = 1.7588 + 0.0761 + 0.6743 + 0.6309 + 0.4352 | **M1** | |
| | 3.58 | **A1** | Accept 3.57 – 3.58 .<br>**SC:** 3.88, if values not combined, scores M1A0. |
| | $H_0$ : distribution fits data<br>N(125, $12^2$) is a good model for the data oe | **B1** | Must mention distribution and data. |
| | 4 degrees of freedom, so tabular value = 9.488<br>3.58 < 9.488<br>Accept $H_0$. | **M1** | Compare their value with 9.488 (9.49) and correct FT conclusion.<br>Or, if values not combined, compare their value with 11.07 and correct FT conclusion for M1A0. |
| | There is sufficient evidence that the normal distribution fits the data<br>There is sufficient evidence that the claim is supported by the data | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | | **6** | |

**Q7.**

| Question | Answer | Marks |
|---|---|---|
| 1 | <table><tr><td>49 E = 56.8</td><td>51 E = 47.6</td><td>40 E = 35.6</td></tr><tr><td>93 E = 85.2</td><td>68 E = 71.4</td><td>49 E = 53.4</td></tr></table> | **M1A1** |
| | $\frac{(49-56.8)^2}{56.8}+\frac{(51-47.6)^2}{47.6}+\frac{(40-35.6)^2}{35.6}+\frac{(93-85.2)^2}{85.2}+\frac{(68-71.4)^2}{71.4}+\frac{(49-53.4)^2}{53.4}$ | **M1** |
| | = 3.096 (3.10) | **A1** |
| | Use appropriate tabular value = 4.605 | **M1** |
| | 3.096 < 4.605 so the grades awarded are independent of the background | **A1** |
| | | **6** |

**Q8.**

| Question | Answer | Marks |
|---|---|---|
| 1 | <table><tr><td>31 E = 22</td><td>35 E = 37.4</td><td>22 E = 28.6</td></tr><tr><td>19 E = 28</td><td>50 E = 47.6</td><td>43 E = 36.4</td></tr></table> | **M1A1** |
| | $\frac{(31-22)^2}{22}+\frac{(35-37.4)^2}{37.4}+\frac{(22-28.6)^2}{28.6}+\frac{(19-28)^2}{28}+\frac{(50-47.6)^2}{47.6}+\frac{(43-36.4)^2}{36.4}$ | **M1** |
| | 9.569 (9.57) | **A1** |
| | Tabular value = 5.991 and comparing 9.569 > 5.991 | **M1** |
| | So standard achieved is dependent on reading scheme used | **A1** |
| | | **6** |

**Q9.**

<table>
<tr><th>Question</th><th>Answer</th><th>Marks</th><th>Guidance</th></tr>
<tr><td>3</td><td>
<table>
<tr><td>$x$</td><td>0</td><td></td><td>2</td><td></td><td></td><td>5</td></tr>
<tr><td>Observed freq</td><td>12</td><td>39</td><td>46</td><td>37</td><td>12</td><td>4</td></tr>
<tr><td>Bin prob</td><td>0.11603</td><td>0.31236</td><td>0.33642</td><td>0.18115</td><td>0.04877</td><td>0.00525</td></tr>
<tr><td>Expected freq</td><td>17.404</td><td>46.858</td><td>50.462</td><td>27.172</td><td>7.316</td><td>0.7878</td></tr>
</table>
</td><td>M1 A1</td><td>Attempt at E values (at least 4 correct)<br><br>All correct, to 2 dp or better</td></tr>
<tr><td></td><td>Add last two columns: 16, 8.104</td><td>M1</td><td></td></tr>
<tr><td></td><td>$\frac{(17.404-12)^2}{17.404}+\frac{(46.858-39)^2}{46.858}+\frac{(50.462-46)^2}{50.462}+$<br>$\frac{(27.172-37)^2}{27.172}+\frac{(8.104-16)^2}{8.104}$ = 14.64 or 14.65</td><td>M1 A1</td><td>Accept 14.6 – 14.7</td></tr>
<tr><td></td><td>’14.64’ > 9.49</td><td>M1</td><td>Compare their value with 9.49</td></tr>
<tr><td></td><td>Freya’s claim is not supported<br>or<br>Data does not fit the distribution</td><td>A1 FT</td><td>Correct conclusion in context, FT only their 14.64<br>SC for Poisson M0M1M1M0 max 2</td></tr>
<tr><td></td><td></td><td>7</td><td></td></tr>
</table>

**Q10.** October/November 2020

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(a) | $\int_0^{0.5}\left(\frac{4}{9}x-\frac{1}{9}x^2\right)\mathrm{d}x \quad \left[=\frac{1}{9}\left[2\times0.5^2-\frac{1}{3}0.5^3\right]\right]$<br><br>$(=0.0509(26)$ or $\frac{11}{216})$ | **M1** | Statement of integration with correct limits |
| | Freq = $p$ = 0.0509(26) × 200 = 10.19 AG | **A1** | Requires sight of 0.0509 or $\frac{11}{216}$ |
| | By addition to 200, $q$ = 26.85 $\left(\frac{725}{27}\right)$ | **B1** | Or by integration |
| | | **3** | |
| 3(b) | <table><tr><td>O</td><td>5</td><td>23</td><td>40</td><td>41</td><td>46</td><td>45</td></tr><tr><td>E</td><td>10.19</td><td>26.85</td><td>37.96</td><td>43.52</td><td>43.52</td><td>37.96</td></tr></table><br>Test statistic = 2.6433 + 0.5520 + 0.1096 + 0.1459 + 0.1413 + 1.3056 | **M1** | At least 4 correct |
| | 4.89 or 4.90 | **A1** | |
| | '4.90' < 11.07: | **M1** | Compare their value with 11.07 |
| | PDF is a satisfactory model for the data. | **A1 FT** | Correct conclusion in context, ft only their 4.90 |
| | | **4** | |

**Q11.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 9 (i) | $E_3 = (3/16) \int_{1.6}^{2.4} (4 - x)^{1/2}\, dx$<br>$= (3/16) [ - (2/3) (4 - x)^{3/2} ]_{1.6}^{2.4}$ | **M1** | State or imply expression for required expected value $E_3$ of $X$ |
| | $= (2 \cdot 4^{3/2} - 1 \cdot 6^{3/2}) / 8 = 1 \cdot 694/8$ *or* $0 \cdot 2118$ | **A1** | Find expected value $E_3$ (may be implied in finding 50 $E_3$)<br>(M1 A1 requires adequate explicit working) |
| | 50 $E_3 = 10 \cdot 59$ AG | **A1** | Hence verify corresponding expected frequency |
| | | **3** | |
| 9(ii) | $_0$: Distribution fits data (AEF) | **B1** | State (at least) null hypothesis in full |
| | $O_i$ 18 16 8 8<br>$E_i$: 14·22 12·54 10·59 12·65 | **M1** | Combine values consistent with all exp. values ≥ 5 |
| | $X^2 = 1 \cdot 005 + 0 \cdot 955 + 0 \cdot 633 + 1 \cdot 709 = 4 \cdot 30$ | **M1 A1** | Find value of $X^2$ from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2/E_i - n$ ] |
| | No. $n$ of cells: 5 4 3<br>$\chi_{n-1,\,0.95}^{\,2}$: 9·488 7·815 5·991 (to 3 s.f.) | **B1** | **FT** on number, $n$, of cells used to find $X^2$<br>State or use consistent tabular value $\chi_{n-1,\,0.95}^{\,2}$ |
| | Accept $H_0$ if $X^2$ < tabular value (AEF) | **M1** | State or imply valid method for conclusion |
| | 4·30 [± 0·01] < 7·81[5] so distribution fits [data]<br>*or* distribution is a suitable model (AEF) | **A1** | Conclusion (requires both values approx. correct) |
| | | **7** | |

**Q12.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 8 | $_0$: Holidays are independent of salesman<br>*or* no association between holidays and salesman (AEF) | **B1** | State (at least) null hypothesis<br><br>Find expected values<br>(lose A1 if rounded to integers)<br>Find value of $X^2$ from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2/E_i - n$ ]<br><br>State or use correct tabular $\chi^2$ value |
| | $E_i$: 29·68 33·04 21·28 (to 1 d.p.)<br>23·32 25·96 16·72 | **M1 A1** | |
| | $X^2$ = 0·7380 + 0·7446 + 0·0037 + 0·9392 + 0·9477 + 0·0047 | **M1** | |
| | = 3·38 (to 3 s.f.) | **A1** | |
| | $\chi_{2,\,0.9}{}^2$ = 4·605 (to 3 s.f.) | **B1** | |
| | Accept $H_0$ if $X^2$ < tabular value | **M1** | Compare their calculated value with their $X^2$ value and appropriate conclusion |
| | Type of holidays is independent of salesman (AEF) | **A1** | Correct conclusion, from correct values $(3.37 - 3.39)$ |
| | | **8** | |

**Q13.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 11B(i) | $\bar{x} = (1/40)\,\Sigma\, x\, f(x) = 68/40 = 1{\cdot}7$ | **B1** | Find mean of sample |
| | $(1/40)\,\Sigma\, x^2\, f(x) = 178/40 = 4{\cdot}45$, Var $= 4{\cdot}45 - 1{\cdot}7^2 = 1{\cdot}56$ | **B1** | Find variance of sample |
| | Mean and variance are similar so Poisson may be suitable | **B1** | State valid comment |
| | | **3** | |
| 11B(ii) | $a = 40 \times 1{\cdot}6^5\, e^{-1{\cdot}6}/5! = 40 \times 0{\cdot}01764$ | **M1** | AG |
| | $a = 0{\cdot}706$ | **A1** | Verify $a$ from Poisson term |
| | $b = 40 - 39{\cdot}758 = 0{\cdot}242$ | **B1*** | Find $b$ |
| | | **3** | |
| 11B(iii) | $H_0$: Distribution fits/models data | **B1** | State (at least) null hypothesis in full |
| | $O_i$: 6 15 9 10<br>$E_i$: 8·076 12·921 10·337 8·666 | **DM1**<br>**A1** | Combine values consistent with all exp. values ⩾ 5<br>(FT on $b$, dep B1*) |
| | $X^2 =$ 0·5337 + 0·3345 + 0·1729 + 0·2053 = 1·25 | **M1** | Find value of $X^2$ from $\Sigma\,(E_i - O_i)^2 / E_i$ [or $\Sigma\, O_i^2/E_i - n$ ] |
| | $X^2 = 1{\cdot}25$ | **A1** | |
| | No. $n$ of cells: 7 6 5 4 3<br>$\chi^2_{n-1,\,0.9}$: 10·64 9·236 7.779 6·251 4·605 | **DB1** | State or use consistent tabular value $\chi^2_{n-1,\,0.9}$ (to 3 s.f.)<br>[FT on number, $n$, of cells used to find $X^2$] |
| | Accept $H_0$ if $X^2$ < tabular value (using their values) | **M1** | AEF |
| | 1·25 [± 0·1] < 6·25 so distn. fits [data]<br>or distn. is a suitable model | **A1** | Conclusion (requires both values approx. correct). AEF |
| | | **8** | |

**Q14.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 8 | $H_0$: Colour is independent of type<br>*or* no association between colour and type (AEF) | **B1** | State (at least) null hypothesis in full |
| | $E_i$: 47·67 43·33 39<br>36·67 33·33 30<br>25·67 23·33 21 (to 3 s.f.) | **M1 A1** | Find expected values $E_i$<br><br>(A0 if rounded to integers) |
| | $X^2 = 0{\cdot}597 + 1{\cdot}241 + 0{\cdot}103$<br>$+ 1{\cdot}603 + 1{\cdot}333 + 0{\cdot}033$<br>$+ \cdot 212 + 0{\cdot}019 + 0{\cdot}429 = 5{\cdot}57$ (to 3 s.f.) | **M1 A1** | Find value of $\chi^2$ from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2 / E_i - n$ ]<br><br>(allow 5·64 for this A1 if integer values of $E_i$ used) |
| | $\chi_{4,\,0.9}^{\,2} = 7{\cdot}779$ *or* 7·78 | **B1** | State or use correct tabular $\chi^2$ value |
| | Accept $H_0$ if $X^2$ < tabular value (AEF) | **M1** | Valid method for reaching conclusion |
| | 5·57 [± 0·01] < 7·78<br>so independent *or* no association (AEF) | **A1** | Correct (abbreviated) conclusion, from approx. correct values |
| | | **8** | |

**Q15.** May/June 2018

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| **11B(i)** | $\bar{x} = 469/250 = 1{\cdot}876$ *and* **AG**<br>$\sigma^2 = (1195 - 469^2/250) / 249 = 1{\cdot}266$ **AG** | **B1** | Verify mean and unbiased variance of data, showing method<br>B0 if biased variance, but allow other marks (max 3/4) |
| | $6 \times 0{\cdot}313 = 1{\cdot}878$ *and* $6 \times 0{\cdot}313 \times 0{\cdot}687 = 1{\cdot}29$[0] | **M1 A1** | Find mean *np and* variance *npq* of binomial distribution |
| | Means and variances are similar for $X$ and B(6, 0·313) | **A1** | Valid explanation (AEF; needs 1·878, 1·290 correct to 0·002) |
| | | **4** | |
| **11B(ii)** | $250 \times {}^6C_4 \times 0{\cdot}313^4 \times 0{\cdot}687^2$ [$= 250 \times 0{\cdot}06795 = 16{\cdot}99$]<br>Allow $49{\cdot}7 \times ({}^6C_4 \times 0{\cdot}313) / ({}^6C_3 \times 0{\cdot}687)$ | **M1 A1** | Show how expected frequency for $x = 4$ is found<br>(AEF for ${}^6C_4 = 15$) |
| | | **2** | |
| **11B(iii)** | $H_0$: [Binomial] distribution fits data (AEF) | **B1** | State (at least) null hypothesis |
| | $O_i$ : 22 83 72 53 20<br>$E_i$ : 26·3 71·9 81·8 49·7 20·3 | ***M1 A1** | Combine last 3 cells so that all exp. value ≥ 5 |
| | $X^2 = 0{\cdot}703 + 1{\cdot}714 + 1{\cdot}174 + 0{\cdot}219 + 0{\cdot}004$<br>$= 3{\cdot}81$ (to 3 s.f.) | **M1 A1** | Find value of $\chi^2$ from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2/E_i - n$ ]<br>(A1 dep *M1) |
| | No. $n$ of cells: 7 6 5<br>$\chi^2_{n-2,\,0.95}$: 11·07 9·488 7·815 | **B1√** | State or use consistent tabular value $\chi^2_{n-2,\,0.95}$ (to 3 s.f.)<br>[FT on number, $n$, of cells used to find $X^2$] |
| | Accept $H_0$ if $X^2 <$ tabular value | **M1** | State or imply valid method for conclusion |
| | $3{\cdot}81$ [± 0·01] $< 7{\cdot}81$[5] so distribution fits<br>*or* scientist's belief is correct | **A1** | Conclusion (AEF; requires both values approx. correct) |
| | | **8** | |

**Q16.**

| | | | |
|---|---|---|---|
| 11B(i) | $\bar{x} = (1/200)\ \Sigma\, x\ \mathrm{f}(x) = 30\,064 / 200 = 150.3$[2] | **M1A1** | Find sample mean to 4 s.f. using mid-interval values<br>(B1 for 29 964 / 200 = 149.8[2] or 30 164 / 200 = 150.8[2]) |
| | | **2** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 11B(ii) | P(151 ⩽ $X$ < 152)<br>= P((151 – 150)/1.2 ⩽ $Z$ < (152 – 150)/1.2)<br>= P(0.833 ⩽ $Z$ < 1.667)<br>= 0.9522 – 0.7975 [= 0.1547] (to 4 s.f.) | **M1A1** | Find P(151 ⩽ $X$ < 152) |
| | $E_{151}$ = 190.44 – 159.5 *or* 200 × 0.1547 = 30.94 AG | **A1** | and hence expected frequency for 151 ⩽ $x$ < 152 |
| | | **3** | |
| 11B(iii) | $H_0$: [Normal] distribution fits data (AEF) | **B1** | State (at least) null hypothesis in full |
| | $O_i$: 3 23 52 69 36 17<br>$E_i$: 9.56 30.94 59.50 59.50 30.94 9.56 | **M1** | Combine values consistent with all exp. values ⩾ 5 |
| | $X^2$ = 4.501 + 2.038 + 0.945 + 1.517 + 0.828 + 5.790 | **M1** | Find value of $X^2$ from $\Sigma\,(E_i - O_i)^2 / E_i$ [or $\Sigma\, O_i^2/E_i - n$ ] |
| | = 15.6 (to 3 s.f.) | **A1** | |
| | No. $n$ of cells: 8 7 6 5<br>$\chi^2_{n-1,\,0.95}$: 14.07 12.59 11.07 9.488 | **B1** | State or use consistent tabular value $\chi^2_{n-1,\,0.95}$ (to 3 s.f.)<br>[**FT** on number, $n$, of cells used to find $X^2$] |
| | Accept $H_1$ if $X^2$ > tabular value (AEF) | **M1** | State or imply valid method for conclusion |
| | 15.6 [± 0.1] > 11.1 so distn. doesn't fit [data] (AEF) | **A1** | Conclusion (requires both values correct) |
| | | **7** | |

**Q17.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 10(i) | $\bar{x} = \Sigma x\, f(x) = 118/40 = 2.95$ AG | **B1** | Verify mean of sample data (B0 for $\bar{x} = 118/40 = 2.95$) |
| | $\sigma_n^2 = (454 - 118^2/40) / 40 = 2.65$<br>*or* $\sigma_{n-1}^2 = (454 - 118^2/40) / 39 = 2.72$ | **B1** | Find variance of sample data (accept either $\sigma_n^2$ or $\sigma_{n-1}^2$ ) |
| | $2.95 \approx 2.65$ o*r* 2.72 | **B1** | Valid comment on correct values |
| | | **3** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 10(ii) | $E_4 = 40\ \lambda^4\ e^{-\lambda}/4!$ with $\lambda = 2.95$ [= 40 × 0.1652]<br>(allow $(\lambda/4) \times E_3 = 0.7375 \times 8.96$) | **M1A1** | State expression for reqd. expected value $E_4$ |
| | | **2** | |
| 10(iii) | $H_0$: [Poisson] distribution fits data (AEF) | **B1** | State (at least) null hypothesis in full |
| | $O_i$: 8 8 10 5 9<br>$E_i$: 8.27 9.11 8.96 6.61 7.05 | **M1** | Combine values consistent with all exp. values ⩾ 5 |
| | $X^2 = 0.0088 + 0.1352 + 0.1207 + 0.3921 + 0.5394$ | **M1** | Find value of $X^2$ from $\Sigma\ (E_i - O_i)^2 / E_i$ [or $\Sigma\ O_i^2/E_i - n$ ] |
| | $= 1.20$ (to 3 s.f.) | **A1** | |
| | No. *n* of cells: 8 7 6 5<br>$\chi^2_{n-2,\,0.95}$: 12.59 11.07 9.488 7.815 | **B1** | State or use consistent tabular value $\chi^2_{n-2,\,0.95}$ (to 3 s.f.)<br>[**FT** on number, *n*, of cells used to find $X^2$] |
| | Accept $H_0$ if $X^2$ < tabular value (AEF) | **M1** | State or imply valid method for conclusion |
| | 1.20 [± 0.1] < 7.81[5] so [Poisson] distn. fits [data]<br>*or* distn. is a suitable model (AEF) | **A1** | Conclusion (requires both values correct) |
| | | **7** | |

**Q18.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 11(b)(i) | $x = (1/250)\ \Sigma\, x\ \mathrm{f}(x) = 414/250 = 1{\cdot}656$ **AG** | **B1** | Verify given mean |
| | **Total:** | **1** | |
| 11(b)(ii) | $p = x/6 = 0{\cdot}276$, $q = 0{\cdot}724$ | **M1 A1** | Use $250\ {}^6\mathrm{C}_i\ q^{6-i}\ p^i$ and find $p$ and $q$ |
| | $a = 250\ {}^6\mathrm{C}_2\ q^4\ p^2 = 78{\cdot}49 \pm 0{\cdot}01$ (to 2 d.p.) | **A2** | Find either exp. value |
| | $b = 250\ {}^6\mathrm{C}_4\ q^2\ p^4 = 11{\cdot}41 \pm 0{\cdot}01$ (to 2 d.p.) | **A1** | Find other exp. value (deduct single **A1** if either value given to only 1 d.p.) |
| | **Total:** | **5** | |
| 11(b)(iii) | $\mathrm{H}_0$: Distribution fits data *or* distribution is binomial (AEF) | **B1** | State (at least) null hypothesis in full<br>Combine values consistent with all exp. values $\geqslant 5$ |
| | $O_i$: 48 69 78 32 23<br>$E_i$: 36·01 82·36 78·49 39·89 13·26 ($\pm$ 0·01) | **M1FT A1** | (FT for **M1** but not **A1** on values of $a$, $b$) |
| | $\chi^2 = 3{\cdot}992 + 2{\cdot}167 + 0{\cdot}003 + 1{\cdot}561 + 7{\cdot}154$ | **M1** | Find $\chi^2$ |
| | $= 14{\cdot}9$ | **A1** | |
| | No. $n$ of cells: 7 6 5 4 3<br>$\chi^2_{n-2,\,0.99}$: 15·09 13·28 11·34 9·210 6·635 | **B1FT** | State or use consistent tabular value $\chi^2_{n-2,\,0.99}$ (to 3 s.f.)<br>[FT on number, $n$, of cells used to find $\chi^2$] |
| | Accept $\mathrm{H}_1$ if $\chi^2 >$ tabular value (AEF)<br>$14{\cdot}9$ [$\pm$ 0·1] > 11·34 so distn. doesn't fit [data] | **M1** | State or imply valid method for conclusion<br>Conclusion (requires both values correct) |
| | *or* manager's belief not justified (AEF) | **A1** | |
| | **Total:** | **8** | |

**Q19.**

| | | | |
|---|---|---|---|
| 10(i) | $\bar{x} = (1/100)\,\Sigma\, x\, f(x) = 325/100 = 3{\cdot}25$ AG | **B1** | Verify given mean (**B0** for $\bar{x} = 325/100 = 3{\cdot}25$) |
| | **Total:** | **1** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 10(ii) | $E_3 = 100\ \lambda^3\ e^{-\lambda}/3!$ with $\lambda = 3{\cdot}25$ | **M1 A1** | State expression for reqd. expected value $E_3$ (**M1** for $E_3$ *or* $E_6$) |
| | $E_6 = 100\ \lambda^6\ e^{-\lambda}/6! = 6{\cdot}35$ | **A1** | Find exp. value $E_6$ |
| | $E_{\geqslant 7} = 100 - \Sigma_0^{\,6}\ E_i = 4{\cdot}77$ | **A1** | Find exp. value $E_{\geqslant 7}$ |
| | **Total:** | **4** | |
| 10(iii) | $H_0$: Distribution fits data (AEF) | **B1** | State (at least) null hypothesis in full |
| | $O_i$: 13 18 26 20 16 7<br>$E_i$: 16·48 20·48 22·18 18·02 11·72 11·12 | **M1FT** | Combine values consistent with all exp. values $\geqslant 5$ (FT on $E_6$ and $E_{\geqslant 7}$) |
| | $\chi^2 = 0{\cdot}735 + 0{\cdot}300 + 0{\cdot}658 + 0{\cdot}218 + 1{\cdot}563 + 1{\cdot}527$ | **M1** | Find $\chi^2$ |
| | $= 5{\cdot}00$ | **A1** | |
| | No. $n$ of cells: 8 7 6 5 4<br>$\chi^2_{n-2,\,0.95}$: 12·59 11·07 9·488 7·815 5·991 | **B1FT** | State or use consistent tabular value $\chi^2_{n-2,\,0.95}$ (to 3 s.f.) [FT on number, $n$, of cells used to find $\chi^2$] |
| | Accept $H_0$ if $\chi^2 <$ tabular value (AEF)<br>5·00 [± 0·1] < 9·49 so distn. fits [data] | **M1** | State or imply valid method for conclusion<br>Conclusion (requires both values correct) |
| | *or* distn. is a suitable model (AEF) | **A1** | |
| | **Total:** | **7** | |

**Q20.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 8 | $_0$: No association<br>*or* method is independent of gender (AEF) | **B1** | State (at least) null hypothesis |
| | $E_i$: 9·0 19·8 31·2<br>6·0 13·2 20·8 | **M1 A1** | Find expected values $E_i$<br>(**A0** if rounded to integers) |
| | $\chi^2 = 0{\cdot}111 + 3{\cdot}073 + 1{\cdot}482$<br>$+ 0{\cdot}167 + 4{\cdot}610 + 2{\cdot}223$ (to 2 d.p.)<br>$= 11{\cdot}7$ (*or* 12·3) | **M1 A1** | Find value of $\chi^2$ from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2/E_i - n$ ]<br>(allow 12·3 if integer values of $E_i$ used) |
| | $\chi_{2,\,0.99}{}^2 = 9{\cdot}21$ | **B1** | State or use correct tabular $\chi^2$ value |
| | Reject $H_0$ if $\chi^2$ > tabular value (AEF) | **M1** | Valid method for reaching conclusion |
| | 11·7 [± 0·1] > 9·21 so there is an association | **A1** | Correct conclusion, from correct values |
| | | **8** | |

**Q21.**

| | | | | |
|---|---|---|---|---|
| **9** | State (at least) null hypothesis (AEF):<br>Find expected values (to 1 d.p.):<br>(lose A1 if one error or rounded to integers)<br>Calculate value of $\chi^2$ (to 3 s.f.):<br><br><br>State or use correct tabular $\chi^2$ value (to 3 s.f.):<br>State or imply valid method for conclusion e.g.:<br>Correct conclusion, from correct values: (AEF) | $H_0$: No difference in preferences<br>21·86718·13329·33310·667<br>19·13315·86725·6679·333<br>$\chi^2 = 0{\cdot}449 + 2{\cdot}075 + 0{\cdot}015 + 0{\cdot}510$<br>$+ 0{\cdot}513 + 2{\cdot}371 + 0{\cdot}017 + 0{\cdot}583$<br>= 6·53 (allow 6·52)<br>$\chi_{3,\,0.9}{}^2$ =6·25[1]<br>Reject $H_0$ if $\chi^2 \geqslant$ tabular value<br>Difference in preferences | **B1**<br><br>**M1 A2**<br><br><br>**M1 A1**<br>**B1**<br>**M1**<br>**A1** | [9] |

**Q22.**

| | | | | |
|---|---|---|---|---|
| **9 (i)** | Find mean and variance of sample data.: | $\bar{x} = 240/60 = 4$ | **B1** | |
| | | $\sigma^2 = 1174/60 - 4^2 = 3{\cdot}57$ | **B1** | |
| | State valid reason why Poisson distn. suitable (AEF): (allow unsuitable since $4 \neq 3{\cdot}57$) | $4 \approx 3{\cdot}57$ (no √ on $\bar{x}, \sigma^2$) | **B1** | [3] |
| **(ii)** | Find expected values $60\lambda^r e^{-\lambda} / r!$ with $\lambda = 4$: | $4{\cdot}40$; $11{\cdot}7_{[2]}$ (to 1 d.p.) | **B1; B1** | [2] |
| **(iii)** | State (at least) null hypothesis (AEF): | $H_0$: [Poisson] distribution fits data | **B1** | |
| | Combine cells so that all exp. value ⩾ 5: | $O_i$: 4 12 … 6 8 | | |
| | | $E_i$: <u>5·50</u> 8·79 … 6·25 <u>6·64</u> | ***M1 A1** | |
| | Calculate value of $\chi^2$ (to 2 d.p.; A1 dep *M1): | $\chi^2 = 0{\cdot}409 + 1{\cdot}172 + 1{\cdot}181$ | | |
| | | $+ 0{\cdot}044 + 1{\cdot}397 + 0{\cdot}81 + 0{\cdot}279$ | | |
| | | $= 5{\cdot}29$ | **M1 DA1** | |
| | State or use consistent tabular value (to 2 d.p.): | 7 cells: $\chi^2_{5,\,0.9} = 9{\cdot}236$ | | |
| | [or fewer or no cells combined: | 8 cells: $\chi^2_{6,\,0.9} = 10{\cdot}64$ | | |
| | | 9 cells: $\chi^2_{7,\,0.9} = 12{\cdot}02$ | | |
| | | 10 cells: $\chi^2_{8,\,0.9} = 13{\cdot}36$ ] | **B1**√ | |
| | State or imply valid method for conclusion e.g.: | Accept $H_0$ if $\chi^2$ < tabular value | **M1** | |
| | Conclusion (AEF, requires both values correct): | $5{\cdot}29 < 9{\cdot}24$ so distn. fits | **A1** | [8] |

**Q23.**

| | | | | | |
|---|---|---|---|---|---|
| **11B** | State suitable distribution: | Geometric | B1 | 1 | |
| | State (at least) null hypothesis: (AEF) | $H_0$: Distn. fits data *or* $p = 0.6$ | B1 | | |
| | (B0 for "It is a good fit") | | | | |
| | Find exp. values using $200pq^{x-1}$ with $p = 0.6$, $q = 0.4$: | 120 48 19.2 7.68 | | | |
| | (ignore incorrect final value here, | 3.072 1.2288 0.8192 | M1 | | |
| | e.g. 0.4915 which can earn max 5/8 | | | | |
| | Combine last 3 cells since exp. value < 5: | *O*: . . . 6 | | | |
| | | *E*: . . . 5.12 | B1 | | |
| | Calculate $\chi^2$ (result correct to 3 s.f.): | $\chi^2 = 0.3 + 0.5208 + 0.4083$ | | | |
| | | $+ 2.8519 + 0.15125 = 4.23$[2] | M1 A1 | | |
| | State or use consistent tabular value (to 3 s.f.): | 5 cells: $\chi_{4,0.95}^2 = 9.488$ | | | |
| | [or if 2 or no cells combined: | 6 cells: $\chi_{5,0.95}^2 = 11.07$ | | | |
| | | 7 cells: $\chi_{6,0.95}^2 = 12.59$ | | | |
| | or if 4 cells combined, as with 0·4915: | 4 cells: $\chi_{3,0.95}^2 = 7.815$ ] | B1 | | |
| | Valid method for reaching conclusion: | Accept $H_0$ if $\chi^2 <$ tabular value | M1 | | |
| | Conclusion (AEF, requires both values correct): | 4.23 < 9.49 so distn fits *or* $p = 0.6$ | A1 | | |
| | (Allow A1 for "It is a good fit") | | | | |
| | Find prob. $p$ of at least one 6 on 5 throws of one die: | $p = 1 - 0.75^5 = 0.7627$ (4 s.f.) | M1 A1 | 8 | |
| | Find prob. of at least one 6 on exactly 4 of 10 dice: | ${}^{10}C_4 \times$ ; $p^4(1-p)^6$ | M1; M1 | | |
| | | $= 210 \times 6.043 \times 10^{-5}$ | | | |
| | | $= 0.0126$ *or* 0.0127 | A1 | 5 | **14** |

**Q24.**

| | | | | | |
|---|---|---|---|---|---|
| **6** | State (at least) null hypothesis (AEF): | $H_0$: Reliability is independent | B1 | | |
| | Find expected values (to 1 d.p.): | 62.253 57.423 | | | |
| | (lose A1 if rounded to integers) | 53.747 49. | M1 A1 | | |
| | Calculate value of $\chi^2$: | $\chi^2 = 0.1212 + 0.5416 + 1.6764$ | | | |
| | | $+ 0.1404 + 0.6274 + 1.941$ | | | |
| | (allow 5.03 if 1 d.p. exp. values used) | $= 5{\cdot}05$ | A1 | | |
| | State or use correct tabular $\chi^2$ value (to 3 s.f.): | $\chi_{2,0.95}^2 = 5.99$[1] | B1 | | |
| | Valid method for reaching conclusion: | Accept $H_0$ if $\chi^2 <$ tabular value | M1 | | |
| | Correct conclusion, from correct values: (AEF) | Reliability is indep. of supplier | A1 | 8 | **8** |

**Q25.**

| | | | | | |
|---|---|---|---|---|---|
| **9** | Find expected frequency $p$: | $p = 200 \int_2^3 (1 / x \ln 8)\,dx$<br>$= (200 / \ln 8)\,[\ln x]_2^3$<br>$= 200 \times 0{\cdot}1950 = 39{\cdot}00$ **A.G.** | M1A1 | | |
| | Find $q$ by similar method *or* by using total of 200: | $q = 21{\cdot}46$ *or* $21{\cdot}45$ | M1 | 4 | |
| | State (at least) null hypothesis: | $H_0$: f ($x$) fits data (A.E.F.) | B1 | | |
| | Calculate $\chi^2$ (to 3 s.f.): | $\chi^2 = 0{\cdot}202 + 0{\cdot}923 + 0{\cdot}678 + 0{\cdot}584$<br>$+ \cdot 134 + 4{\cdot}134 + 3{\cdot}644 = 11{\cdot}3$ | M1A | | |
| | State or use correct tabular $\chi^2$ value (to 3 s.f.):: | $\chi^2_{6,\,0.95} = 12{\cdot}59$ | B1 | | |
| | Valid method for reaching conclusion: | Accept $H_0$ if $\chi^2 \leqslant$ tabular value | M1 | | |
| | Conclusion consistent with correct values (A.E.F): | Distribution fits observations | A1 | 6 | **10** |

# 4. Non-Parametric Tests

**Q1.**

**6** A teacher at a large college gave a mathematical puzzle to all the students. The median time taken by a random sample of 24 students to complete the puzzle was 18.0 minutes. The students were then given practice in solving puzzles. Two weeks later, the students were given another mathematical puzzle of the same type as the first. The times, in minutes, taken by the random sample of 24 students to complete this puzzle are as follows.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 18.2 | 17.5 | 16.4 | 15.1 | 20.5 | 26.5 | 19.2 | 23.2 |
| 17.9 | 18.8 | 25.8 | 19.9 | 17.7 | 16.2 | 17.3 | 16.6 |
| 17.1 | 20.1 | 20.3 | 12.6 | 16.0 | 21.4 | 22.7 | 18.4 |

The teacher claims that the practice has not made any difference to the average time taken to complete a puzzle of this type.

Carry out a Wilcoxon signed-rank test, at the 10% significance level, to test whether there is sufficient evidence to reject the teacher's claim. [10]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2022

**Q2.**

**5** A manager claims that the lengths of the rubber tubes that his company produces have a median of 5.50 cm. The lengths, in cm, of a random sample of 11 tubes produced by this company are as follows.

5.56 5.45 5.47 5.58 5.54 5.52 5.60 5.35 5.59 5.51 5.62

It is required to test at the 10% significance level the null hypothesis that the population median length is 5.50 cm against the alternative hypothesis that the population median length is not equal to 5.50 cm.

Show that both a sign test and a Wilcoxon signed-rank test give the same conclusion and state this conclusion. [9]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q3.**

**5** Georgio has designed two new uniforms $X$ and $Y$ for the employees of an airline company. A random sample of 11 employees are each asked to assess each of the two uniforms for practicality and appearance, and to give a total score out of 100. The scores are given in the table.

| Employee | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* | *K* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Uniform $X$ | 82 | 74 | 42 | 59 | 60 | 73 | 94 | 98 | 62 | 36 | 50 |
| Uniform $Y$ | 78 | 75 | 63 | 56 | 67 | 82 | 99 | 90 | 72 | 48 | 61 |

**(a)** Give a reason why a Wilcoxon signed-rank test may be more appropriate than a $t$-test for investigating whether there is any evidence of a preference for one of the uniforms. [1]

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Carry out a Wilcoxon matched-pairs signed-rank test at the 10% significance level. [7]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

**Q4.**

**2** A company is developing a new flavour of chocolate by varying the quantities of the ingredients. A random selection of 9 flavours of chocolate are judged by two tasters who each give marks out of 100 to each flavour of chocolate.

| Chocolate | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* |
|---|---|---|---|---|---|---|---|---|---|
| Taster 1 | 72 | 86 | 75 | 92 | 98 | 79 | 87 | 60 | 62 |
| Taster 2 | 84 | 72 | 74 | 95 | 85 | 87 | 82 | 75 | 68 |

Carry out a Wilcoxon matched-pairs signed-rank test at the 10% significance level to investigate whether, on average, there is a difference between marks awarded by the two tasters. [7]

**Q5.**

**6** The blood cholesterol levels, measured in suitable units, of a random sample of 11 women and a random sample of 12 men are shown below.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Women | 51 | 55 | 242 | 167 | 152 | 256 | 75 | 137 | 98 | 238 | 235 | |
| Men | 311 | 262 | 170 | 302 | 175 | 320 | 220 | 260 | 72 | 351 | 86 | 333 |

Carry out a Wilcoxon rank-sum test, at the 5% significance level, to test whether, on average, there is a difference in cholesterol levels between women and men. [9]

© UCLES 2021

**Q6.**

**4** Applicants for a particular college take a written test when they attend for interview. There are two different written tests, *A* and *B*, and each applicant takes one or the other. The interviewer wants to determine whether the medians of the distribution of marks obtained in the two tests are equal. The marks obtained by a random sample of 8 applicants who took test *A* and a random sample of 8 applicants who took test *B* are as follows.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Test *A* | 46 | 32 | 29 | 12 | 33 | 18 | 25 | 40 |
| Test *B* | 36 | 28 | 49 | 37 | 48 | 35 | 41 | 31 |

**(a)** Carry out a Wilcoxon rank-sum test at the 5% significance level to determine whether there is a difference in the population median marks obtained in the two tests. [6]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

The interviewer considers using the given information to carry out a paired sample *t*-test to determine whether there is a difference in the population means for the two tests.

**(b)** Give two reasons why it is not appropriate to use this test. [2]

**Q7.**

**2** The times, in milliseconds, taken by a computer to perform a certain task were recorded on 10 randomly chosen occasions. The times were as follows.

6.44 6.16 5.62 5.82 6.51 6.62 6.19 6.42 6.34 6.28

It is claimed that the median time to complete the task is 6.4 milliseconds.

**(a)** Carry out a Wilcoxon signed-rank test at the 5% significance level to test this claim. [6]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** State an underlying assumption that is made when using a Wilcoxon signed-rank test. [1]

.....

.....

.....

.....

.....

**Q8.**

**6** A biologist is studying the effect of nutrients on the heights to which plants grow. A random sample of 24 similar young plants is divided into two equal groups *A* and *B*. The plants in group *A* are fed with nutrients and water and the plants in group *B* are given only water. After four weeks, the height, in cm, of each plant is measured and the results are as follows.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group *A* | 12.3 | 11.8 | 12.1 | 13.2 | 11.1 | 10.6 | 13.8 | 12.0 | 12.2 | 12.4 | 13.5 | 13.9 |
| Group *B* | 11.7 | 10.8 | 10.9 | 11.3 | 11.2 | 12.6 | 11.0 | 10.5 | 11.9 | 12.5 | 10.7 | 11.6 |

The biologist decides to carry out a test at the 5% significance level to test whether the nutrients have resulted in an increase in growth.

**(a)** She carries out a Wilcoxon rank-sum test. Give a reason why this is an appropriate choice of test. [1]

.....

.....

.....

.....

.....

.....

**(b)** Carry out the Wilcoxon rank-sum test for these results. [10]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2020

© UCLES 2020

**Q9.**

**2** Metal rods produced by a certain factory are claimed to have a median breaking strength of 200 tonnes. For a random sample of 9 rods, the breaking strengths, measured in tonnes, were as follows.

210 186 188 208 184 191 215 198 196

A scientist believes that the median breaking strength of metal rods produced by this factory is less than 200 tonnes.

**(a)** Use a Wilcoxon signed-rank test, at the 5% significance level, to test whether there is evidence to support the scientist's belief. [6]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(b)** Give a reason why a Wilcoxon signed-rank test is preferable to a sign test, when both are valid. [1]

.....

.....

.....

.....

**Q10.**

**2** A large school is holding an essay competition and each student has submitted an essay. To ensure fairness, each essay is given a mark out of 100 by two different judges. The marks awarded to the essays submitted by a random sample of 12 students are shown in the following table.

| Student | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* | *K* | *L* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Judge 1 | 62 | 74 | 52 | 48 | 68 | 55 | 56 | 64 | 37 | 70 | 81 | 59 |
| Judge 2 | 65 | 70 | 47 | 49 | 76 | 74 | 67 | 54 | 50 | 77 | 72 | 75 |

**(a)** One of the students claims that Judge 2 is awarding higher marks than Judge 1.

Carry out a Wilcoxon matched-pairs signed-rank test at the 5% significance level to test whether the data supports the student's claim. [7]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

It is discovered later that the marks awarded to student *A* have been entered incorrectly. In fact, Judge 1 awarded 65 marks and Judge 2 awarded 62 marks.

**(b)** By considering how this change affects the test statistic, explain why the conclusion of the test carried out in part **(a)** remains the same. [2]

.....

.....

.....

.....

.....

# MS
# Non-Parametric Tests

**Q1.**

May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6 | Differences/rank:<br>0.2 \| – 0.5 \| –1.6 \| –2.9 \| 2.5 \| 8.5 \| 1.2 \| 5.2<br>**2 \| –5 \| –11 \| –18 \| 17 \| 24 \| 9 \| 21**<br>–0.1 \| 0.8 \| 7.8 \| 1.9 \| –0.3 \| –1.8 \| –0.7 \| –1.4<br>**–1 \| 7 \| 2 \| 13 \| –3 \| –12 \| –6 \| –10**<br>–0.9 \| 2.1 \| 2.3 \| –5.4 \| –2.0 \| 3.4 \| 4.7 \| 0.4<br>**–8 \| 15 \| 16 \| –22 \| –14 \| 19 \| 20 \| 4** | **M1 A1** | Attempt at differences, allow 4 errors. |
| | Q = 110 P = 190 | **A1** | Either sum correct. |
| | $H_0$: population median =18.0 and $H_1$: population median $\neq$ 18.0 | **B1** | Must be 'population', allow *m*. |
| | Mean = $\frac{1}{4}n(n+1) = \frac{1}{4}\times 24\times 25 = 150$ | **B1** | Normal approximation: mean. |
| | Variance = $\frac{1}{24}n(n+1)(2n+1) = \frac{1}{24}\times 24\times 25\times 49 = 1225$ | **B1** | |
| | $\frac{110.5-150}{\sqrt{1225}}$ | **M1** | Attempt at test statistic, allow incorrect or missing cc, FT *their* 110. |
| | $-1.129$ | **A1** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6 | 10% 2-tail, critical value is 1.645<br>$-1.129 > -1.645$ or 0.1296 > 0.05, accept $H_0$. | **M1** | Compare with $-1.645$, correct FT conclusion. |
| | Insufficient evidence to suggest that median is not 18.0/<br>Insufficient evidence to reject teacher's claim. | **A1** | Correct conclusion, in context.<br>Level of uncertainty in language is used. |
| | | **10** | |

**Q2.** May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5 | Signs: + – – + + + + – + + + and use Binomial | **M1** | 11 signs (values not needed) allow one error. |
| | Consider B(11, 0.5) with P( $X \geqslant 8$) or P( $X \leqslant 3$) | **M1** | Use of correct binomial. |
| | Compare 0.227 with 0.1 or 0.113 with 0.05 | **A1** | |
| | Differences:<br>0.06 –0.05 –0.03 0.08 0.04 0.02<br>0.10 –0.15 0.09 0.01 0.12 | **M1** | Allow one error. |
| | Signed Ranks:<br>6 -5 -3 7 4 2<br>9 -11 8 1 10 | **M1** | Allow one swap. |
| | (P = 47) Q = 19 | **A1** | |
| | Critical value from table = 13 | **B1** | |
| | 19 > 13<br>Accept $H_0$ | **M1** | FT *their* (19) compared with 13 and correct FT conclusion |
| | Insufficient evidence to support median not being 5.50 | **A1** | BOTH conclusions correct from correct working, in context.<br>Level of uncertainty in language. |
| | | **9** | |

**Q3.** May/June 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5(a) | Underlying distribution or population of differences is unknown<br>Underlying distributions or population of scores for both $X$ and $Y$ are unknown | **B1** | Not known to be normal.<br>Condone 'scores for $X$ and $Y$ cannot be assumed to be normally distributed'. |
| | | **1** | |
| 5(b) | $H_0$ : difference of population medians = 0 | **B1** | Correct hypotheses, allow $m$ but not $\mu$ or mean. |
| | $H_1$ : difference of population median $\neq 0$ | **B1** | 'population' included. |
| | Diff: 4 −1 −21 3 −7 −9 −5 8 −10 −12 −11 | **M1** | Differences, allow one error. |
| | Rank: 3 −1 −11 2 −5 −7 −4 6 −8 −10 −9 | **M1** | Signed ranks, allow one sign error. |
| | $[Q = -55,]\, P = 11$ | **A1** | $P = 11$ |
| | Critical tabular value = 13<br>'11' < 13 so reject $H_0$ | **M1** | Comparison with 13 and correct ft conclusion. |
| | Sufficient evidence of a preference for one of the uniforms OE | **A1** | Correct conclusion, in context, following correct work, except possibly second B1.<br>Level of uncertainty in language is used. |
| | | **7** | |

**Q4.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2 | $_0$: difference in population medians = 0 ($m_a = m_b$) | **B1** | Correct hypotheses, allow $m_d$ . Do not allow $\mu$ or mean. |
| | $H_1$: difference in population medians ≠ 0 ($m_a \neq m_b$) | **B1** | 'population' included. |
| | Diff: 12 −14 −1 3 −13 8 −5 15 6 | **M1** | Allow one error. |
| | Rank: 6 −8 −1 2 −7 5 −3 9 4 | **M1** | Allow one error. |
| | Smaller sum 19 | **A1** | (Other sum is 26.) |
| | Critical tabular value = 8<br>'19' > 8 so accept $H_0$ | **M1** | Comparison with 8 and correct FT conclusion. |
| | There is insufficient evidence that marks differ | **A1** | Correct conclusion, in context, following correct work, except possibly the hypotheses. Level of uncertainty in language is used. |
| | | **7** | |

**Q5.** October/November 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6 | <table><tr><td>51</td><td>1</td><td>72</td><td>3</td></tr><tr><td>55</td><td>2</td><td>86</td><td>5</td></tr><tr><td>75</td><td>4</td><td>170</td><td>10</td></tr><tr><td>98</td><td>6</td><td>175</td><td>11</td></tr><tr><td>137</td><td>7</td><td>220</td><td>12</td></tr><tr><td>152</td><td>8</td><td>260</td><td>17</td></tr><tr><td>167</td><td>9</td><td>262</td><td>18</td></tr><tr><td>235</td><td>13</td><td>302</td><td>19</td></tr><tr><td>238</td><td>14</td><td>311</td><td>20</td></tr><tr><td>242</td><td>15</td><td>320</td><td>21</td></tr><tr><td>256</td><td>16</td><td>333</td><td>22</td></tr><tr><td></td><td></td><td>351</td><td>23</td></tr></table> | **M1** | Attempt at ranking. |
| | Total ranks: 95 | **A1** | |
| | $H_0 : m_x = m_y$ and $H_1 : m_x \neq m_y$ | **B1** | Allow words, must include ‘population’ median. |
| | Use normal approximation with attempts at mean and variance. | **M1** | |
| | Mean =132, variance = 264 | **A1** | |
| | $\frac{95.5 - 132}{\sqrt{264}}$ | **M1** | Allow no or wrong continuity correction for M1 only. |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6 | – 2.246 | **A1** | CAO |
| | Critical value is – 1.96.<br>– 2.246 < – 1.96 reject $H_0$. | **M1** | Compare their value with – 1.96.<br>Or area comparison 0.0123 or 0.0124 with 0.025 and FT conclusion. |
| | There is sufficient evidence of a difference in levels. | **A1** | Correct conclusion, in context, following correct work.<br>Level of uncertainty in language is used. |
| | | **9** | |

**Q6.**



| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4(a) | <table><tr><td>12</td><td>1</td><td>28</td><td>4</td></tr><tr><td>18</td><td>2</td><td>31</td><td>6</td></tr><tr><td>25</td><td>3</td><td>35</td><td>9</td></tr><tr><td>29</td><td>5</td><td>36</td><td>10</td></tr><tr><td>32</td><td>7</td><td>37</td><td>11</td></tr><tr><td>33</td><td>8</td><td>41</td><td>13</td></tr><tr><td>40</td><td>12</td><td>48</td><td>15</td></tr><tr><td>46</td><td>14</td><td>49</td><td>16</td></tr></table> | **M1** | Attempt at ranking. |
| | Test statistic: 52 | **A1** | |
| | $H_0 : m_x = m_y$ and $H_1 : m_x \neq m_y$ | **B1** | Allow in words but 'population' must be included. |
| | Critical value for (8, 8) is 49. | ***B1** | Allow 51 if clearly one-tail test in hypotheses. |
| | 52 > 49<br>Accept $H_0$ | **DM1** | Compare their calculated value with 49 and correct FT conclusion. |
| | Insufficient evidence of difference in medians. | **A1** | Correct conclusion, in context, following correct work. Level of uncertainty in language is used. |
| | | **6** | |
| 4(b) | Not a paired sample. | **B1** | |
| | Underlying distribution/population not (known to be) normal<br>Underlying distribution/population unknown | **B1** | B0 for 'data is not normally distributed'<br>B0 for 'marks are not normally distributed' |
| | | **2** | |

**Q7.**

| Question | Answer | Marks |
|---|---|---|
| 2(a) | $_0$: population median is 6.40, $H_1$: population median ≠ 6.40 | **B1** |
| | Calculate differences and signed ranks<br>0.04 –0.24 –0.78 –0.58 0.11 0.22 –0.21 0.02 –0.06 –0.12<br>2 –8 –10 –9 4 7 –6 1 –3 –5 | **M1A1** |
| | Test statistic = 4 + 7 + 2 +1 =14 | **A1** |
| | Compare with correct critical value 8 | **M1** |
| | Accept $H_0$: insufficient evidence to reject claim (**FT** *their* test statistic) | **A1 FT** |
| | | **6** |
| 2(b) | Symmetrically distributed about the median | **B1** |
| | | **1** |

**Q8.**

| Question | Answer | Marks |
|---|---|---|
| 6(a) | Difference of location test for populations not known to be normal | **B1** |
| | | **1** |

| | | |
|---|---|---|
| 6(b) | | **M1** |

| | | | |
|---|---|---|---|
| 10.6 | 2 | 10.5 | 1 |
| 11.1 | 7 | 10.7 | 3 |
| 11.8 | 12 | 10.8 | 4 |
| 12.0 | 14 | 10.9 | 5 |
| 12.1 | 15 | 11.0 | 6 |
| 12.2 | 16 | 11.2 | 8 |
| 12.3 | 17 | 11.3 | 9 |
| 12.4 | 18 | 11.6 | 10 |
| 13.2 | 21 | 11.7 | 11 |
| 13.5 | 22 | 11.9 | 13 |
| 13.8 | 23 | 12.5 | 19 |
| 13.9 | 24 | 12.6 | 20 |

| | |
|---|---|
| Total ranks: 109 or 191 | **A1** |
| $H_0$: $m_x = m_y$ $H_1$: $m_x > m_y$ | **B1** |
| Use normal approximation with attempts at mean and variance | **M1** |
| mean = 150, variance = 300 | **A1** |

| | | |
|---|---|---|
| 6(b) | $\frac{109.5-150}{\sqrt{300}} = -2.338$ | **M1A1** |
| | Probability 0.0097 | **A1** |
| | This is less than 0.05, so reject $H_0$ | **M1** |
| | There is evidence of an effect on growth | **A1** |
| | | **10** |

**Q9.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2(a) | $_0$: $m = 200$ $H_1$: $m < 200$ | **B1** | Allow 'median' in words, allow $m$ not defined |
| | 10 −14 −12 8 −16 −9 15 −2 −4 | **M1** | Signed differences (at most 3 errors) |
| | Ranks:<br>5 −7 −6 3 −9 −4 8 −1 −2 | **A1** | Award for correct rank order, ignore signs |
| | Sum ranks T = 16 | **A1** | CWO |
| | Critical value 8 and compare '16' > 8 | **M1** | Compare their T with 8 |
| | Accept $H_0$<br>Insufficient evidence to support scientist's belief | **A1** | In context, all correct, except possibly hypotheses.<br>Level of uncertainty in language used. No contradictions |
| | | **6** | |
| 2(b) | Magnitude of differences from median are taken into account | **B1** | Must mention magnitude and differences |
| | | **1** | |

**Q10.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2(a) | $_0$: difference in (population) medians = 0<br>$H_1$: difference in (population) medians < 0 (or >0) | **B1 B1** | Correct hypotheses (median or *m*)<br>All notation identified e.g. *m* = population median if used |
| | Differences:<br>–3 4 5 –1 –8 –19 –11 10 –13 –7 9 –16 | **M1** | At most 3 errors |
| | Ranks:<br>–2 3 4 –1 –6 –12 –9 8 –10 –5 7 –11 | **A1** | Award for correct rank order, ignore signs |
| | Sum T = 3 + 4 + 8 + 7 = 22 | **A1** | cwo |
| | Compare with critical value 17: '22' > 17 | **M1** | Compare *their* T with 17 |
| | Accept $H_0$<br>Data does not support student's claim | **A1** | In context, all correct, except possibly second B1<br>Level of uncertainty in language used. No contradictions. |
| | | **7** | |
| 2(b) | Rank for A becomes + 2, | **M1** | T = 24 |
| | Changing sign of difference can only reduce evidence in favour of the claim. | **A1** | still > 17 and test result unchanged |
| | | **2** | |

# 5. Probability Generating Functions

**Q1.**

**2** The probability generating function, $G_Y(t)$, of the random variable $Y$ is given by

$$G_Y(t) = 0.04 + 0.2t + 0.37t^2 + 0.3t^3 + 0.09t^4.$$

**(a)** Find Var($Y$). [4]

The random variable $Y$ is the sum of two independent observations of the random variable $X$.

**(b)** Find the probability generating function of $X$, giving your answer as a polynomial in $t$. [3]

**Q2.**

**3** George throws two coins, $A$ and $B$, at the same time. Coin $A$ is biased so that the probability of obtaining a head is $a$. Coin $B$ is biased so that the probability of obtaining a head is $b$, where $b < a$. The probability generating function of $X$, the number of heads obtained by George, is $G_X(t)$. The coefficients of $t$ and $t^2$ in $G_X(t)$ are $\frac{5}{12}$ and $\frac{1}{12}$ respectively.

**(a)** Find the value of $a$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable $Y$ is the sum of two independent observations of $X$.

**(b)** Find the probability generating function of $Y$, giving your answer as a polynomial in $t$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2022

.....

.....

.....

.....

.....

.....

.....

.....

**(c)** Find Var($Y$). [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q3.**

**6** Tanji has a bag containing 4 red balls and 2 blue balls. He selects 3 balls at random from the bag, without replacement. The number of red balls selected by Tanji is denoted by $X$.

**(a)** Find the probability generating function $G_X(t)$ of $X$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Tanji also has two coins, each biased so that the probability of obtaining a head when it is thrown is $\frac{1}{4}$. He throws the two coins at the same time. The number of heads obtained is denoted by $Y$.

**(b)** Find the probability generating function $G_Y(t)$ of $Y$. [2]

.....

.....

.....

.....

.....

.....

.....

The random variable $Z$ is the sum of the number of red balls selected by Tanji and the number of heads obtained.

**(c)** Find the probability generating function of $Z$, expressing your answer as a polynomial. [3]

.....

.....

.....

.....

© UCLES 2021

**(d)** Use the probability generating function of $Z$ to find $\mathrm{E}(Z)$ and $\mathrm{Var}(Z)$. [5]

**Q4.**

**4** $X$ is a discrete random variable which takes the values 0, 2, 4, … . The probability generating function of $X$ is given by

$$G_X(t) = \frac{1}{3-2t^2}.$$

**(a)** Find $E(X)$ and $\text{Var}(X)$. [5]

**(b)** Find $P(X = 4)$. [3]

**Q5.**

**5** Nine balls labelled 1, 2, 3, 4, 5, 6, 7, 8, 9 are placed in a bag. Kai selects three balls at random from the bag, without replacement. The random variable $X$ is the number of balls selected by Kai that are labelled with a multiple of 3.

**(a)** Find the probability generating function $G_X(t)$ of $X$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The balls are replaced in the bag.

Jacob now selects two balls at random from the bag, without replacement. The random variable $Y$ is the number of balls selected by Jacob that are labelled with an even number.

**(b)** Find the probability generating function $G_Y(t)$ of $Y$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

The random variable $Z$ is the sum of the number of balls that are labelled with a multiple of 3 selected by Kai and the number of balls that are labelled with an even number selected by Jacob.

**(c)** Find the probability generating function of $Z$, expressing your answer as a polynomial. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(d)** Use the probability generating function of $Z$ to find $\mathrm{E}(Z)$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**Q6.**

**5** The random variable $X$ is such that $\mathrm{P}(X = r) = kr^2$ for $r = 1, 2, 3, 4,$ where $k$ is a constant.

**(a)** Find the value of $k$. [1]

.....

.....

.....

.....

.....

.....

**(b)** Find the probability generating function $\mathrm{G}_X(t)$ of $X$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable $Y$ has probability generating function $\mathrm{G}_Y(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$.

The random variable $Z$ is the sum of $X$ and $Y$.

**(c)** Assuming that $X$ and $Y$ are independent, find the probability generating function $\mathrm{G}_Z(t)$ of $Z$ as a polynomial in $t$. [3]

.....

.....

.....

.....

.....

.....

.....

© UCLES 2021

**(d)** Given that $\mathrm{E}(Z) = \frac{13}{3}$, use $\mathrm{G}_Z(t)$ to find $\mathrm{Var}(Z)$. [3]

**Q7.**

**6** A bag contains 4 red balls and 6 blue balls. Rassa selects two balls at random, without replacement, from the bag. The number of red balls selected by Rassa is denoted by $X$.

**(a)** Find the probability generating function, $G_X(t)$, of $X$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

Rassa also tosses two coins. One coin is biased so that the probability of a head is $\frac{2}{3}$. The other coin is biased so that the probability of a head is $p$. The probability generating function of $Y$, the number of heads obtained by Rassa, is $G_Y(t)$. The coefficient of $t$ in $G_Y(t)$ is $\frac{7}{12}$.

**(b)** Find $G_Y(t)$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2020

The random variable $Z$ is the sum of the number of red balls selected and the number of heads obtained by Rassa.

**(c)** Find the probability generating function of $Z$, expressing your answer as a polynomial. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(d)** Use the probability generating function of $Z$ to find E($Z$). [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2020

**Q8.**

**4** The discrete random variable $X$ has probability generating function $G_X(t)$ given by

$$G_X(t) = 0.2t + 0.5t^2 + 0.3t^3.$$

The random variable $Y$ is the sum of two independent observations of $X$.

**(a)** Find the probability generating function of $Y$, giving your answer as an expanded polynomial in $t$. [3]

**(b)** Use the probability generating function of $Y$ to find $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$. [5]

**Q9.**

**5** Keira has two unbiased coins. She tosses both coins. The number of heads obtained by Keira is denoted by $X$.

**(a)** Find the probability generating function $G_X(t)$ of $X$. [1]

.....

.....

.....

.....

.....

.....

.....

Hassan has three coins, two of which are biased so that the probability of obtaining a head when the coin is tossed is $\frac{1}{3}$. The corresponding probability for the third coin is $\frac{1}{4}$. The number of heads obtained by Hassan when he tosses these three coins is denoted by $Y$.

**(b)** Find the probability generating function $G_Y(t)$ of $Y$. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

The random variable $Z$ is the total number of heads obtained by Keira and Hassan.

**(c)** Find the probability generating function of $Z$, expressing your answer as a polynomial. [3]

.....

.....

.....

.....

© UCLES 2020

.....

.....

.....

.....

.....

.....

**(d)** Use the probability generating function of $Z$ to find $\mathrm{E}(Z)$. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**(e)** Use the probability generating function of $Z$ to find the most probable value of $Z$. [1]

.....

.....

.....

.....

.....

.....

**Q10.**

**5** The random variable $X$ has the binomial distribution B($n$, $p$).

**(a)** Write down an expression for $\mathrm{P}(X = r)$ and hence show that the probability generating function of $X$ is $(q + pt)^n$, where $q = 1 - p$. [3]

**(b)** Use the probability generating function of $X$ to prove that $\mathrm{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$. [5]

# MS
# Probability Generating Function

**Q1.** May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 2(a) | $G_Y'(t) = 0.2 + 0.74t + 0.9t^2 + 0.36t^3$<br>$G_Y''(t) = 0.74 + 1.8t + 1.08t^2$ | **M1** | PGF differentiated twice. |
| | $G_Y'(1) = 2.2$<br>$G_Y''(1) = 3.62$ | **A1** | |
| | $\text{Var}(Y) = G_Y''(1) + G_Y'(1) - (G_Y'(1))^2$<br>$3.62 + 2.2 - 2.2^2$ | **M1** | Correct formula used and attempt at Var($Y$). |
| | 0.98 | **A1** | |
| | | **4** | |
| 2(b) | $G_Y(t) = \left(a + bt + ct^2\right)^2$ | **M1*** | Correct method. |
| | So $a^2 = 0.04, \quad a = 0.2$<br>$c^2 = 0.09, \quad c = 0.3$<br>$2ab = 0.2, \quad b = \frac{0.1}{0.2} = 0.5$ | **depM1** | Attempt to find all coefficients. |
| | $G_X(t) = 0.2 + 0.5t + 0.3t^2$ | **A1** | *Note:* $G_Y = \frac{1}{100}(1+t)^2(2+3t)^2$ M1, accept without $\frac{1}{100}$.<br>So $G_X = \frac{1}{10}(1+t)(2+3t)$ dep M1, must have $\frac{1}{10}$.<br>Final answer A1. |
| | | **3** | |

**Q2.** May/June 2022

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 3(a) | $G_X(t) = \frac{6}{12} + \frac{5}{12}t + \frac{1}{12}t^2$<br>P(0 heads) = $(1-a)(1-b) = \frac{6}{12}$<br>P(1 head) = $a(1-b) + (1-a)b = \frac{5}{12}$<br>P(2 heads) = $ab = \frac{1}{12}$<br>$\left(a + b = \frac{7}{12},\ ab = \frac{1}{12}\right)$ | **M1** | Obtain two correct equations and attempt to solve. |
| | Solve to give<br>$a = \frac{1}{3}$ $\left(b = \frac{1}{4}\right)$ | **A1** | Correct value for $a$. |
| | | **2** | |
| 3(b) | $G_Y(t) = \left(\frac{6}{12} + \frac{5}{12}t + \frac{1}{12}t^2\right)^2$ | **M1** | Square their $G_X(t)$. |
| | $\frac{1}{144}\left(36 + 60t + 37t^2 + 10t^3 + t^4\right)$ | **M1** | Obtain quartic polynomial from 3-term $G_X(t)$. |
| | | **A1** | |
| | | **3** | |
| 3(c) | $G_Y'(t) = \frac{1}{144}\left(60 + 74t + 30t^2 + 4t^3\right)$<br>$G_Y''(t) = \frac{1}{144}\left(74 + 60t + 12t^2\right)$ | **M1** | Differentiate their $G_Y(t)$ twice. |
| | Var($Y$) = $G_Y''(1) + G_Y'(1) - (G_Y'(1))^2$<br>$\frac{1}{144}(146) + \frac{7}{6} - \frac{49}{36} = \frac{59}{72} = 0.819$ | **M1** | Use correct formula and attempt at Var($Y$). |
| | 0.819 | **A1** | |

**Q3.** May/June 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(a) | P(3R, 2R, 1R, 0R) = $\frac{1}{5},\frac{3}{5},\frac{1}{5},0$ | **M1** | At least 2 correct probabilities used in a polynomial. |
| | $G_X(t)=\frac{1}{5}t+\frac{3}{5}t^2+\frac{1}{5}t^3$ | **A1** | |
| | | **2** | |
| 6(b) | P(2H, 1H, 0H) = $\frac{1}{16},\frac{6}{16},\frac{9}{16}$ | **M1** | One correct probability and all three adding to 1 used in a polynomial. |
| | $G_Y(t)=\frac{9}{16}+\frac{6}{16}t+\frac{1}{16}t^2$ | **A1** | |
| | | **2** | |
| 6(c) | Attempt to multiply results from part **(a)** and part **(b)** | **M1** | |
| | Obtain polynomial | **M1** | |
| | $\frac{9}{80}t+\frac{33}{80}t^2+\frac{28}{80}t^3+\frac{9}{80}t^4+\frac{1}{80}t^5$ | **A1** | Accept exact equivalent decimals. |
| | | **3** | |
| 6(d) | E($Z$) = $G_Z'(1)=\frac{1}{80}(9+66+84+36+5)$ | **M1** | Differentiate and put $t = 1$. |
| | $=\frac{200}{80}=2.5$ | **A1** | |
| | $G_Z''(1)=\frac{1}{80}(66+168+108+20)=4.525=\left[\frac{362}{80}\right]$ | **M1** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 6(d) | Var $(Z) = G''(1) + G'(1) - \left(G'(1)\right)^2$ | **M1** | Use result. |
| | $\frac{362}{80} + \frac{200}{80} - \left(\frac{200}{80}\right)^2 = 0.775 = \left[\frac{31}{40}\right]$ | **A1** | |
| | | **5** | |

**Q4.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 4(a) | G'($t$) = $\frac{4t}{\left(3-2t^2\right)^2}$ | **M1** | Differentiate to obtain $\frac{kt}{\left(3-2t^2\right)^2}$ OE. |
| | so E(X) = G'(1) = 4 | **A1** | CAO WWW |
| | G''($t$) = $\frac{12+24t^2}{\left(3-2t^2\right)^3}$ or $4\left(3-2t^2\right)^{-2} + 32t^2\left(3-2t^2\right)^{-3}$ | **M1** | OE. Differentiate, allow only numerical or sign slips. |
| | Var($X$) = G''(1) + G'(1) – $\left(\text{G}'(1)\right)^2$ = 36 + 4 – 16 | **M1** | Substitute their values into correct formula, dependent on attempt at G''($t$). |
| | [Var($X$)] = 24 | **A1** | CAO WWW |
| | | **5** | |
| 4(b) | $G_X(t) = \frac{1}{3-2t^2} = \left(3-2t^2\right)^{-1}$<br>$= \frac{1}{3}\left(1+\frac{2}{3}t^2+\frac{4}{9}t^4+\ldots.\right)$ | **M1** | Expand given expression or give expression for the term in $t^4$ . |
| | P($X$ = 4) = their coefficient of $t^4$ | **M1** | Found from legitimate method. |
| | [P($X$ = 4) =] $\frac{4}{27}$ | **A1** | WWW |
| | | **3** | |

**Q5.**

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5(a) | P(3, 6, 9) = $\frac{3}{9}\times\frac{2}{8}\times\frac{1}{7}=\frac{6}{504}$<br>P(Two of 3, 6, 9) = $\frac{3}{9}\times\frac{2}{8}\times\frac{6}{7}\times 3=\frac{108}{504}$<br>**P(o** 6, **9)** $\frac{3}{9}\times\frac{6}{8}\times\frac{5}{7}\times 3=\frac{270}{504}$<br>**P(no** 3, **6,** $\frac{6}{9}\times\frac{5}{8}\times\frac{4}{7}=\frac{120}{504}$ | **B1** | At least 2 probabilities correct. |
| | $G_X(t)=\frac{20}{84}+\frac{45}{84}t+\frac{18}{84}t^2+\frac{1}{84}t^3$ | **M1**<br>**A1** | Attempt with at least 3 probabilities in a polynomial, CAO. |
| | | **3** | |
| 5(b) | P(both even ) = $\frac{12}{72}$<br>P(one even) = $\frac{40}{72}$<br>P(no even) = $\frac{20}{72}$ | **M1** | |
| | $G_Y(t)=\frac{5}{18}+\frac{10}{18}t+\frac{3}{18}t^2$ | **A1** | CAO |
| | | **2** | |

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5(c) | $\frac{1}{1512}\left(\frac{20}{84}+\frac{45}{84}t+\frac{18}{84}t^2+\frac{1}{84}t^3\right)\left(\frac{5}{18}+\frac{10}{18}t+\frac{3}{18}t^2\right)$ | **M1** | Method and attempt to multiply. |
| | | **M1** | Multiplication to obtain single quintic polynomial. |
| | $\frac{1}{1512}(100+425t+600t^2+320t^3+64t^4+3t^5)$ | **A1** | CAO |
| | | **3** | |
| 5(d) | $G'(1)=\frac{1}{1512}(425+1200+960+256+15)$ | **M1** | |
| | $\frac{17}{9}$ | **A1** | CWO |
| | | **2** | |

**Q6.** October/November 2021

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5(a) | $(1+4+9+16)k=1$ so $k=\frac{1}{30}$ | **B1** | |
| | | **1** | |
| 5(b) | $G_X(t)=\frac{1}{30}t+\frac{4}{30}t^2+\frac{9}{30}t^3+\frac{16}{30}t^4$ | **M1A1** | Using their $k$ in a polynomial, at least two terms correct for their $k$. |
| | | **2** | |
| 5(c) | $\left(\frac{1}{30}t+\frac{4}{30}t^2+\frac{9}{30}t^3+\frac{16}{30}t^4\right)\left(\frac{1}{4}+\frac{1}{2}t+\frac{1}{4}t^2\right)$ | **M1** | Method and attempt to multiply. |
| | $\frac{1}{120}\left(t+6t^2+18t^3+38t^4+41t^5+16t^6\right)$ | **M1A1** | Multiplication to obtain single polynomial of order 6. |
| | | **3** | |
| 5(d) | Given: G'(1) = 13/3<br>G''($t$) = $\frac{1}{120}(12+108t+456t^2+820t^3+480t^4)$ | **M1** | Differentiate twice. |
| | Var($X$) = G''(1) + $\frac{13}{3}-\left(\frac{13}{3}\right)^2$ | **M1** | Use correct formula. |
| | $\frac{1876}{120}+\frac{13}{3}-\frac{169}{9}=\frac{107}{90}$ or 1.19 | **A1** | CAO |
| | | **3** | |

**Q7.** May/June 2020

| Question | Answer | Marks |
|---|---|---|
| 6(a) | P(BB) = $\frac{6}{10}\times\frac{5}{9}=\frac{1}{3}$ P(RB, BR) = $\frac{6}{10}\times\frac{4}{9}=\frac{8}{15}$ P(RR) = $\frac{4}{10}\times\frac{3}{9}=\frac{2}{15}$ | **M1** |
| | $\frac{1}{3}+\frac{8}{15}t+\frac{2}{15}t^2$ (**FT** *their* probabilities) | **A1 FT** |
| | | **2** |
| 6(b) | P(1 H) = $\frac{2}{3}(1-p)+\frac{1}{3}p=\frac{7}{12}$ | **M1** |
| | $p=\frac{1}{4}$ | **A1** |
| | PGF of $Y=\frac{1}{4}+\frac{7}{12}t+\frac{1}{6}t^2$ | **A1** |
| | | **3** |
| 6(c) | PGF of $Z=\left(\frac{1}{3}+\frac{8}{15}t+\frac{2}{15}t^2\right)\left(\frac{1}{4}+\frac{7}{12}t+\frac{1}{6}t^2\right)$ | **B1M1** |
| | $=\frac{1}{180}\left(15+59t+72t^2+30t^3+4t^4\right)$ **AEF** | **A1** |
| | | **3** |
| 6(d) | Attempt to differentiate their $G_Z(t)$ and evaluate $G'_Z(1)$ | **M1** |
| | E($Z$) = (59 + 144 + 90 + 16)/180 = $\frac{309}{180}=1.72$ | **A1** |
| | | **2** |

**Q8.**

| Question | Answer | Marks |
|---|---|---|
| 4(a) | $G_X(t) = 0.2t + 0.5t^2 + 0.3t^3$ | **B1** |
| | $G_Y(t) = \left(0.2t + 0.5t^2 + 0.3t^3\right)^2$ | **M1** |
| | $= 0.04t^2 + 0.2t^3 + 0.37t^4 + 0.3t^5 + 0.09t^6$ | **A1** |
| | | **3** |
| 4(b) | $G'_Y(t) = 0.08t + 0.6t^2 + 1.48t^3 + 1.5t^4 + 0.54t^5$ | **M1** |
| | $E(Y) = 4.2$ | **A1** |
| | $G''_Y(t) = 0.08 + 1.2t + 4.44t^2 + 6t^3 + 2.7t^4$ | **M1** |
| | Use $G''_Y(1) + G'_Y(1) - (G'_Y(1))^2$ | **M1** |
| | $Var(Y) = 14.42 + 4.2 - 4.2^2 = 0.98$ | **A1** |
| | | **5** |

**Q9.** October/November 2020

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5(a) | $G_X(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$ | **B1** | Accept $(0.5+0.5t)^2$ |
| | | **1** | |
| 5(b) | P(0H)= $\frac{12}{36}$ P(1H) = $\frac{16}{36}$ P(2H) = $\frac{7}{36}$ P(3H) = $\frac{1}{36}$ | **M1 A1** | Attempt at probs, at least 2 correct<br>All correct |
| | $G_Y(t) = \frac{12}{36} + \frac{16}{36}t + \frac{7}{36}t^2 + \frac{1}{36}t^3$ | **B1 FT** | FT *their* probabilities, must be cubic with 4 non-zero terms |
| | | **3** | |
| 5(c) | $G_Z(t) = (\frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2)\left(\frac{12}{36} + \frac{16}{36}t + \frac{7}{36}t^2 + \frac{1}{36}t^3\right)$ | **M1** | Attempt to multiply their two PGF |
| | $= \frac{1}{144}\left(12 + 40t + 51t^2 + 31t^3 + 9t^4 + t^5\right)$ | **M1 A1** | Obtain quintic expression and collect terms |
| | | **3** | |
| 5(d) | $G'_Z(t) = \frac{1}{144}\left(40 + 102t + 93t^2 + 36t^3 + 5t^4\right)$ | **M1** | Differentiate |
| | $E(Z) = G'_Z(1) = \frac{23}{12} = (= 1.92)$ | **A1** | |
| | | **2** | |
| 5(e) | 2 | **B1 FT** | FT power of term with largest coefficient in *their* $G_Z(t)$ |
| | | **1** | |

**Q10.** October/November 2020

| Question | Answer | Marks | Guidance |
|---|---|---|---|
| 5(a) | $P(X = r) = {}^{n}C_r p^r (1\text{-}p)^{n-r}$ or ${}^{n}C_r p^r q^{n-r}$ | **B1** | |
| | $G_X(t) = \sum_{0}^{n} nCr\, p^r (1-p)^{n-r} t^r$ | **M1** | Accept minimum of 4 terms including the last.<br>Shown with specific value of $n$ is M0 |
| | $\sum_{0}^{n} nCr (pt)^r (1-p)^{n-r}$<br>$= (q + pt)^n$ | **A1** | At least one intermediate step to be shown, with $p$ and $t$ grouped<br><br>AG |
| | | **3** | |
| 5(b) | $G'_X(t) = n(q + pt)^{n-1} \times p$ | **M1** | |
| | So $E(X) = G'_X(1) = np(q + p)^{n-1}$ and $q + p = 1$<br>so $E(X) = np$ | **A1** | |
| | $G''_X(t) = n(n-1)(q + pt)^{n-2} \times p \times p$ | **M1** | |
| | $Var(X) = n(n-1)p^2 + np - (np)^2$ | **M1** | |
| | $np(1-p)$ | **A1** | |
| | | **5** | |